

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M nucleic - nucleic search, using sw model

Run on: September 11, 2004, 10:06:53 ; Search time 14891.6 Seconds
(without alignments)
7040.653 Million cell updates/sec

Title: US-09-939-408a-12

Perfect score: 2419

Sequence: 1 ctccctgcctcactcctaaccc.....gagttcaaaaaaaaaa 2419

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb ba:*
2: gb hcg:*
3: gb in:*
4: gb om:*
5: gb ov:*
6: gb pat:*
7: gb pl:*
8: gb pr:*
9: gb ro:*
10: gb scs:*
11: gb sy:*
12: gb un:*
13: gb vi:*
14: gb vi:*
15: em ba:*
16: em fun:*
17: em hum:*
18: em in:*
19: em ml:*
20: em om:*
21: em or:*
22: em ov:*
23: em pat:*
24: em ph:*
25: em pl:*
26: em ro:*
27: em sts:*
28: em un:*
29: em vi:*
30: em hcg_hum:*
31: em hcg_inv:*
32: em hcg_other:*
33: em hcg_mus:*
34: em hcg_pln:*
35: em hcg_rod:*
36: em hcg_mam:*
37: em hcg_vrt:*
38: em sy:*
39: em hcg_hum:*
40: em hcg_mus:*
41: em hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2418.6	100.0	2419	6	AR199602
2	2418.6	100.0	2419	6	AX366860
3	1826.6	75.5	2741	6	AX366876
4	1629	67.3	2475	6	AR199605
5	1629	67.3	2475	6	AX366868
6	1196.6	49.5	2151	6	E01785
7	1196.6	49.5	2151	6	AR205032
8	1196.6	49.5	2151	6	AR282639
9	1196.6	49.5	2151	6	AX082544
10	1196.6	49.5	2151	6	AX370659
11	1196.6	49.5	2315	8	RTPALGNA
12	1196.6	49.5	2331	6	E01783
13	1196.6	49.5	2331	6	E01784
14	1196.6	49.5	2331	6	E01785
15	1196.6	49.5	2439	6	AR199605
16	1196.6	49.5	2439	6	AX366866
17	1195	49.4	2151	6	E01543
18	1195	49.4	2151	6	AR205031
19	1195	49.4	2151	6	AR282638
20	1195	49.4	2151	6	AX082542
21	1195	49.4	2151	6	AX370657
22	1188.8	49.1	2061	6	AR282658
23	1188.8	49.1	2061	6	AX370681
24	1136.6	47.0	2787	6	AR199604
25	1136.6	47.0	2787	6	AX366864
26	802.4	33.2	2986	8	RHPAL2
27	802.4	33.2	3340	8	RHPAL2
28	747.2	30.9	3586	8	RHPAL2
29	402	16.6	2124	6	AX653389
30	402	16.6	2124	6	AX660330
31	402	16.6	2510	8	AK100346
32	401.4	16.6	2015	8	AY224546
33	399.8	16.5	2488	6	MZEPAL
34	399.8	16.5	2126	6	AX660228
35	393.2	16.3	139226	8	OSUNO274
36	393.8	16.3	139488	8	OSUNO262
37	388.8	16.1	170186	2	AC135419
38	388.8	16.1	177374	8	AC135429
39	385.8	15.9	2504	8	AK102817
40	384.2	15.9	2211	6	AX660248
41	384.2	15.9	2280	6	AX658420
42	381.4	15.8	2282	6	AX660254
43	379.8	15.7	2594	8	AK120127
44	379.4	15.7	2157	6	AX654419
45	377.8	15.6	2108	6	AX660252

ALIGNMENTS

RESULT 1
AR199602
LOCUS AR199602 2419 bp DNA
DEFINITION Sequence 12 from patent US 6355468.
ACCESSION AR199602
VERSION AR199602.1 GI:20249676
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2419)
Yoshida,R.K. and Kootstra,A.B.
Phenylalanine ammonia lyase polypeptide and polynucleotide
sequences and methods of obtaining and using same
JOURNAL Patent: US 6355468-A 12 12-MAR-2002;

FEATURES
source

Location/Qualifiers
1..2419
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 2418.6; DB 6; Length 2419;
Best Local Similarity 100.0%; Pred. No. 1e-284;
Matches 2418; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGAGCTCACTTAACCCGCTACTAGACTCCGAGAGGCCCTTCCTTGAATCCGCTC 60
DB 1 CTCGAGCTCACTTAACCCGCTACTAGACTCCGAGAGGCCCTTCCTTGAATCCGCTC 60
QY 61 GCCACACGCTCGCCAAAGGCTTTTACCAAGGCTCGACGCGCTCCGACCAATCGGCT 120
DB 61 GCCACACGCTCGCCAAAGGCTTTTACCAAGGCTCGACGCGCTCCGACCAATCGGCT 120
QY 121 GCGGGCCCACTTCGAGCTTCGCGCGCAAGCCGAGCTTCGATGCGCACGCCGCGACAG 180
DB 121 GCGGGCCCACTTCGAGCTTCGCGCGCAAGCCGAGCTTCGATGCGCACGCCGCGACAG 180
QY 181 TCGAGACTCGAATCGTGAAGAGCTCTCAAGCAAGCCCAAGCAAGCTTCGAGCTC 240
DB 181 TCGAGACTCGAATCGTGAAGAGCTCTCAAGCAAGCCCAAGCAAGCTTCGAGCTC 240
QY 241 AGCGGATACAGCTTACCGTCCGTAAGCTTCGCGCGCGCGCGCAAGGCGCAGGCTC 300
DB 241 AGCGGATACAGCTTACCGTCCGTAAGCTTCGCGCGCGCGCGCGCAAGGCGCAGGCTC 300
QY 301 CGGATCGAAGAGAGAGAGATCCGCGCAAGCGCTCGAAGAGAGCTTCGATTCGATAG 360
DB 301 CGGATCGAAGAGAGAGAGATCCGCGCAAGCGCTCGAAGAGAGCTTCGATTCGATAG 360
QY 361 GCCCACTTCAGAACTCGGCTTCAAGAGCTCAAGAGGCTTCGATGCGCGCGCGCACAG 420
DB 361 GCCCACTTCAGAACTCGGCTTCAAGAGCTCAAGAGGCTTCGATGCGCGCGCGCACAG 420
QY 421 AAGACTGAGAGATGAGAGTCAAGCTCCAGAAAGGCGCTCAATCGAGCACAGCTTCGCGG 480
DB 421 AAGACTGAGAGATGAGAGTCAAGCTCCAGAAAGGCGCTCAATCGAGCACAGCTTCGCGG 480
QY 481 AGCGGAGCTCCGATCGCTTCGCTTCAAGCTTCGAGAGCGCGCTTCGAGAAACAGCTTCG 540
DB 481 AGCGGAGCTCCGATCGCTTCGCTTCAAGCTTCGAGAGCGCGCTTCGAGAAACAGCTTCG 540
QY 541 GAGGTGCTCCGCGCGCATGATCCGCTCAACTGCTCAAGCTTCGAGCGCTTCGCGC 600
DB 541 GAGGTGCTCCGCGCGCATGATCCGCTCAACTGCTCAAGCTTCGAGCGCTTCGCGC 600
QY 601 GTCGCGCTTCGCTTCGAGAGCGCTCAAGCTTCGAGAGCGCGCTTCGATTCGATGCG 660
DB 601 GTCGCGCTTCGCTTCGAGAGCGCTCAAGCTTCGAGAGCGCGCTTCGATTCGATGCG 660
QY 661 GTCCCTTCGCGCGCTTCGATTCGAGAGCGCGCTTCGAGAGCGCGCTTCGATTCGATG 720
DB 661 GTCCCTTCGCGCGCTTCGATTCGAGAGCGCGCTTCGAGAGCGCGCTTCGATTCGATG 720
QY 721 GCGCGCATCAAGGCTTCAGAGCTTCAGAGCTTCAGAGCTTCAGAGCTTCAGAGCTTC 780
DB 721 GCGCGCATCAAGGCTTCAGAGCTTCAGAGCTTCAGAGCTTCAGAGCTTCAGAGCTTC 780
QY 781 ATCATGTTTGGCGGAGGCTTCGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAG 840
DB 781 ATCATGTTTGGCGGAGGCTTCGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAG 840
QY 841 AAGAGAGCTTCGCTTCGATTCAGAGAGCGCGCTTCGCGCTTCGATTCGAGAGCTTCG 900
DB 841 AAGAGAGCTTCGCTTCGATTCAGAGAGCGCGCTTCGCGCTTCGATTCGAGAGCTTCG 900
QY 901 CTGCAAGCTTCGAGAGCTTCGCTTCGAGAGCGCGCTTCGAGAGCTTCGAGAGCTTCG 960
DB 901 CTGCAAGCTTCGAGAGCTTCGCTTCGAGAGCGCGCTTCGAGAGCTTCGAGAGCTTCG 960

QY 961 GCATGCTCGCGCAGAGAGGCTTCGCGCGCTTCAGAGAGCTTCGCGCGCGCAG 1020
DB 961 GCATGCTCGCGCAGAGAGGCTTCGCGCGCTTCAGAGAGCTTCGCGCGCGCAG 1020
QY 1021 CCGGCGAGCTCGAGGCTTCGCGCGCATTCGAGAGCTTCGCGCGCTTCGCGCTTCGCG 1080
DB 1021 CCGGCGAGCTCGAGGCTTCGCGCGCATTCGAGAGCTTCGCGCGCTTCGCGCTTCGCG 1080
QY 1081 GTTGAAGCAGAGAGAGGCTTCAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCT 1140
DB 1081 GTTGAAGCAGAGAGAGGCTTCAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCT 1140
QY 1141 TACCGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAG 1200
DB 1141 TACCGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAGAGCTTCGAG 1200
QY 1201 TACTGAGCTTCGCTTCGAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCTTC 1260
DB 1201 TACTGAGCTTCGCTTCGAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCTTC 1260
QY 1261 AACAGAGAGAGAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAG 1320
DB 1261 AACAGAGAGAGAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAG 1320
QY 1321 AAGAGAGAGCTTCGAGCTTCGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCT 1380
DB 1321 AAGAGAGAGCTTCGAGCTTCGAGAGCTTCAGAGAGCTTCAGAGAGCTTCAGAGAGCT 1380
QY 1381 CTCAAGCTTCGAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAG 1440
DB 1381 CTCAAGCTTCGAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAG 1440
QY 1441 AACTATACAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAG 1500
DB 1441 AACTATACAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAG 1500
QY 1501 CTTGCGAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAG 1560
DB 1501 CTTGCGAGAGAGAGAGCTTCGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAG 1560
QY 1561 TCGCTGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1620
DB 1561 TCGCTGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1620
QY 1621 CTGCTGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1680
DB 1621 CTGCTGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1680
QY 1681 TTCAAGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAG 1740
DB 1681 TTCAAGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAG 1740
QY 1741 CTGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1800
DB 1741 CTGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1800
QY 1801 ACAGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAG 1860
DB 1801 ACAGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAG 1860
QY 1861 GTGCTGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1920
DB 1861 GTGCTGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1920
QY 1921 TGAAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1980
DB 1921 TGAAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTC 1980
QY 1981 TGGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCT 2040
DB 1981 TGGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCTTCGAGAGAGCT 2040

[illegible]

Query Match	100.0%	Score 2418.6	DB 6	Length 2419
Best Local Similarity	100.0%	Pred. No. 1e-284		
Matches 2419	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	CTCTGACCTCACTTAACCCCGTCACTAGCACTCGACATGAGGCCCTTCTTGATCTCCTC	60	
Db	1	CTCTTGTCTCACTTAACCCCGTCACTAGCACTCGACATGAGGCCCTTCTTGATCTCCTC	60	
QY	61	GCCACCAAGCTCGCCCAACGGCTTTACCAAGCGCTCGACAGCGCGCTCCGACCAAGTCGACT	120	
Db	61	GCCACCAAGCTCGCCCAACGGCTTTACCAAGCGCTCGACAGCGCGCTCCGACCAAGTCGACT	120	
QY	121	GCGGAGCCCAATTGGCGCTTCGCGCGACAGCGCGCGCTTCGATGGCCACAGCGCGCACACAG	180	
Db	121	GCGGAGCCCAATTGGCGCTTCGCGCGACAGCGCGCGCTTCGATGGCCACAGCGCGCACACAG	180	
QY	181	TCGCAAGCTCGAGATCGTGCAGAGAGCTCTTCACGACACCCACGACGACGTCGTGAGCTC	240	
Db	181	TCGCAAGCTCGAGATCGTGCAGAGAGCTCTTCACGACACCCACGACGACGTCGTGAGCTC	240	
QY	241	AGCGGGTTCAGCTTCACCGTCCGTTGAGTGTGTGGGCGCGCGCCGACAGAGGGCGAGGGCTC	300	
Db	241	AGCGGGTTCAGCTTCACCGTCCGTTGAGTGTGTGGGCGCGCGCCGACAGAGGGCGAGGGCTC	300	
QY	301	CGCGTCCAGAACGACGACGAGATCCGCGCACGCGCTCGACAAAGAGCGTTCGACTTCCTCAAG	360	
Db	301	CGCGTCCAGAACGACGACGAGATCCGCGCACGCGCTCGACAAAGAGCGTTCGACTTCCTCAAG	360	
QY	361	GCCCAAGCTTCAGAACTCGGTTACGAGATCACACAGGGTTTCGGTGGCTCGGCGCACAGC	420	
Db	361	GCCCAAGCTTCAGAACTCGGTTACGAGATCACACAGGGTTTCGGTGGCTCGGCGCACAGC	420	
QY	421	AGGACTGAGATGAGTACAGTACAGCTCCAGAGGCGCTCATTCGACGACAGCTTCGGGCGTG	480	
Db	421	AGGACTGAGATGAGTACAGTACAGCTCCAGAGGCGCTCATTCGACGACAGCTTCGGGCGTG	480	
QY	481	ACGCGCAAGTCCGCTCTCGTCTTACGCGTGGAGCGGCGCTTCGAGAACAGGCTTCCGCTC	540	
Db	481	ACGCGCAAGTCCGCTCTCGTCTTACGCGTGGAGCGGCGCTTCGAGAACAGGCTTCCGCTC	540	
QY	541	GAGGTCTCGCGCGCGCGCGCGCATAGTATCCGCGTCACTCGCTCAGCGCGGCGCATCTCGGCC	600	
Db	541	GAGGTCTCGCGCGCGCGCGCGCATAGTATCCGCGTCACTCGCTCAGCGCGGCGCATCTCGGCC	600	
QY	601	GTCGCGCTCGCTGCTCTTGAAGGCGCTCACCACTTCTTGAACCAACCGCATCACGCCCATC	660	
Db	601	GTCGCGCTCGCTGCTCTTGAAGGCGCTCACCACTTCTTGAACCAACCGCATCACGCCCATC	660	
QY	661	GTCCCGCTCGCGGCTCCATTCGCGCGTGGGCGACTCAAGCCGCTTCGTACATCGCC	720	
Db	661	GTCCCGCTCGCGGCTCCATTCGCGCGTGGGCGACTCAAGCCGCTTCGTACATCGCC	720	
QY	721	GAGCGCATCACCGGTCACCGCGAGTCAAGTTCACGTTTGGACGAGGAGACGAGAG	780	
Db	721	GAGCGCATCACCGGTCACCGCGAGTCAAGTTCACGTTTGGACGAGGAGACGAGAG	780	
QY	781	ATCATGTTTGGCGCGAGGCGCATTCGCTCTTGGTGTGAGGCACTGTCCTCGGCGCG	840	
Db	781	ATCATGTTTGGCGCGAGGCGCATTCGCTCTTGGTGTGAGGCACTGTCCTCGGCGCG	840	
QY	841	AAGGAGGCTTCGAGTCGAGTCAAGGAAAGGCGCGCTTCGAGTGGGAGACCTTCAGT	900	
Db	841	AAGGAGGCTTCGAGTCGAGTCAAGGAAAGGCGCGCTTCGAGTGGGAGACCTTCAGT	900	
QY	901	CTGCACGACTCGACATGCTCTGCTCTTCGACAGGCTTGAAGGCTCTCACGATGAG	960	
Db	901	CTGCACGACTCGACATGCTCTGCTCTTCGACAGGCTTGAAGGCTCTCACGATGAG	960	
QY	961	GCGATGCTCGGCGACAGAGGCTCGTTCGCGCGCTTACATCACGACGTCGTGGCGCGCAC	1020	
Db	961	GCGATGCTCGGCGACAGAGGCTCGTTCGCGCGCTTACATCACGACGTCGTGGCGCGCAC	1020	

QY	102	CCGGGACAGTGGAGGTGGCGGCAACA	TCGCAAGCTCTCTTCGCGCTCGTGGTTGAC	1089
Dp	1021	CCGGGACAGTGGAGGTGGCGGCAACA	TCGCAAGCTCTCTTCGCGCTCGTGGTTGAC	1080
QY	1081	GTTGACACGAGAGAGAGGTCAAGGTCA	AGGACACAGGCGCATTTCTTCGCCAGGACCGC	1140
Dp	1081	GTTGACACGAGAGAGAGGTCAAGGTCA	AGGACACAGGCGCATTTCTTCGCCAGGACCGC	1140
QY	1141	TACCCGCTCCGACGTGGCTTCAAGTTCT	CGGCCCCGTGTTGAGAGACATGATGACGCC	1200
Dp	1141	TACCCGCTCCGACGTGGCTTCAAGTTCT	CGGCCCCGTGTTGAGAGACATGATGACGCC	1200
QY	1201	TACTGACTCTCTGCTGAGAAACAACA	CGACACGACACACCCTCTTCACAGCTCGAG	1260
Dp	1201	TACTGACTCTCTGCTGAGAAACAACA	CGACACGACACACCCTCTTCACAGCTCGAG	1260
QY	1261	AACAGCAGACCGGCGACGGCGGCACTT	CCAGCGTGTCTGTATTTGATGAGAG	1320
Dp	1261	AACAGCAGACCGGCGACGGCGGCACTT	CCAGCGTGTCTGTATTTGATGAGAG	1320
QY	1321	AACAGCAGGCTGGACTGGCCCTCATGGG	AAAGTCAACTCAACGAGAGCACCGAGTGG	1380
Dp	1321	AACAGCAGGCTGGACTGGCCCTCATGGG	AAAGTCAACTCAACGAGAGCACCGAGTGG	1380
QY	1381	CTGACGCTGCATTAACCGCGGCGCTGT	GTGCTTCGCTTCGATCCAGAGACCCGTGCT	1440
Dp	1381	CTGACGCTGCATTAACCGCGGCGCTGT	GTGCTTCGCTTCGATCCAGAGACCCGTGCT	1440
QY	1441	AACATTCAGAGGCAAGGGCTTGGACAT	TCAATCGCTGTCTTAAGCTTCGAGCTCGG	1500
Dp	1441	AACATTCAGAGGCAAGGGCTTGGACAT	TCAATCGCTGTCTTAAGCTTCGAGCTCGG	1500
QY	1501	CTTGCCAAACCGGCTCACTACCTTGTGT	CGAGCGGACAGATGGGCAACCAAGCCGTCA	1560
Dp	1501	CTTGCCAAACCGGCTCACTACCTTGTGT	CGAGCGGACAGATGGGCAACCAAGCCGTCA	1560
QY	1561	TGCGTCGCTTCATTTCCGCGCGCGCAT	TCGCGAGGCGAACGACGCTCTTCTCTCTT	1620
Dp	1561	TGCGTCGCTTCATTTCCGCGCGCGCAT	TCGCGAGGCGAACGACGCTCTTCTCTCTT	1620
QY	1621	CTGCGCTGCACCTGTACCTGACGCTCA	AGGCGCTCAAGCTTCGCGGATGGAGGTGAC	1680
Dp	1621	CTGCGCTGCACCTGTACCTGACGCTCA	AGGCGCTCAAGCTTCGCGGATGGAGGTGAC	1680
QY	1681	TTTCAGAGACATTTGACACCGCTTCTCC	GCATCTCTCCAGAGACCTCGGACCTGGC	1740
Dp	1681	TTTCAGAGACATTTGACACCGCTTCTCC	GCATCTCTCCAGAGACCTCGGACCTGGC	1740
QY	1741	CTTGAGGTCAAGGCACTTGCGCTGAGGT	CAAGAAGCGCTCAACAAGCTTTCGAGAG	1800
Dp	1741	CTTGAGGTCAAGGCACTTGCGCTGAGGT	CAAGAAGCGCTCAACAAGCTTTCGAGAG	1800
QY	1801	ACGACAGACTAGACTTCGAGCCGCGCTG	CAAGACTTCGTCGTCGACGCGGAC	1860
Dp	1801	ACGACAGACTAGACTTCGAGCCGCGCTG	CAAGACTTCGTCGTCGACGCGGAC	1860
QY	1861	GTCGTGAGACTCTCTGTGTCTGCGCTT	CTGCGAAGTCAACCTTACTGCGGTCAACGG	1920
Dp	1861	GTCGTGAGACTCTCTGTGTCTGCGCTT	CTGCGAAGTCAACCTTACTGCGGTCAACGG	1920
QY	1921	TGGAGAGTTCCTGGCGCGAAGAGCATCT	CGCTCAACGCGAGAGTGGCAACCGCTTC	1980
Dp	1921	TGGAGAGTTCCTGGCGCGAAGAGCATCT	CGCTCAACGCGAGAGTGGCAACCGCTTC	1980
QY	1981	TGGCAGACGCGCTTTGCGAGCGCGCGAC	GCATACCTTCTGCGCGGACGCGGCTC	2040
Dp	1981	TGGCAGACGCGCTTTGCGAGCGCGCGAC	GCATACCTTCTGCGCGGACGCGGCTC	2040
QY	2041	CTGTACTCTGTTGTCGAGAGACTCGAGT	GAAGCGCGCGGCGGCGACGATGTTTGTG	2100
Dp	2041	CTGTACTCTGTTGTCGAGAGACTCGAGT	GAAGCGCGCGGCGGCGACGATGTTTGTG	2100

Qy	210	GGGGTGCAAGCAAGAACATCGSAGCAAGCTCCGCGCATCTACAGGCGCATCAAGAC	2160
Db	2101	GGGGTGCAAGCAAGAACATCGSAGCAAGCTCCGCGCATCTACAGGCGCATCAAGAC	2160
Qy	2161	GGCGCGATCAACCAAGTCTCTGTAAATGCTGCGTAAGCCCGACAAAGCTTGCTTA	2220
Db	2161	GGCGCGATCAACCAAGTCTCTGTAAATGCTGCGTAAGCCCGACAAAGCTTGCTTA	2220
Qy	2221	GAAGCGCGCGCTACCCCAAGACAAGCTTTTGAAGCTGTGATGCGCAAGACGAGCTTT	2280
Db	2221	GAAGCGCGCGCTACCCCAAGACAAGCTTTTGAAGCTGTGATGCGCAAGACGAGCTTT	2280
Qy	2281	CCTCATACACATGTGCGCTTACTCTCTGCCGCTCATACAGTCTCAAGTTTTCGAT	2340
Db	2281	CCTCATACACATGTGCGCTTACTCTCTGCCGCTCATACAGTCTCAAGTTTTCGAT	2340
Qy	2341	CCCGGCTCTCTCGGTGCTCAAGTACACAGCTGTATAGAGCCTGGAATGGATTCAAGTTCG	2400
Db	2341	CCCGGCTCTCTCGGTGCTCAAGTACACAGCTGTATAGAGCCTGGAATGGATTCAAGTTCG	2400
Qy	2401	AGTTCAAAAAAAAAAAAAA 2419	
Db	2401	AGTTCAAAAAAAAAAAAAA 2419	

RESULT 3	AX366876	2741 bp	DNA	linear	PAT 15-FEB-2002
LOCUS	AX366876				
DEFINITION	Sequence 28 from Patent WO0208402.				
ACCESSION	AX366876				
VERSION	AX366876.1	GI:18698164			
KEYWORDS					
SOURCE	Rhodotorula graminis				
ORGANISM	Rhodotorula graminis				
	Eukaryota; Fungi; Basidiomycota; Urediniomycetes;				
	Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;				
	Rhodotorula.				
REFERENCE	1				
AUTHORS	Yoshida, R.K. and Koeststra, A.B.				
TITLE	Phenylalanine ammonia lyase polypeptide and polynucleotide				
	sequences and methods of obtaining and using same				
JOURNAL	Patent: WO 0208402-A 28 31-JAN-2002;				
	PCBU Services, Inc. (US)				
FEATURES					
source	1..2741				
	Location/Qualifiers				
	/organism="Rhodotorula graminis"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:29898"				
exon	1..361				
intron	362..448				
exon	449..880				
intron	881..960				
exon	961..1295				
intron	1296..1354				
exon	1355..1529				
intron	1530..1586				
exon	1587..1748				
intron	1749..1821				
exon	1822..1947				
intron	1948..2007				
exon	2008..2586				
terminator	2587..2589				
ORIGIN					
Query Match	75.5%;	Score 1826.6;	DB 6;	Length 2741;	
Best Local Similarity	84.3%;	Pred. No. 7.4e-213;			
Matches 2312;	Conservative	2;	Mismatches	1;	Indels 426; Gaps 6;
QY	37	ATGGGCGCTTCCTTGAGCTGAGCTGGCGACAGAGCTGGCCAAAGGCTTACCAAGGCGCTG	96		
Db	1	ATGGGCGCTTCCTTGAGCTGAGCTGGCGACAGAGCTGGCCAAAGGCTTACCAAGGCGCTG	60		
Y	97	CACGGCGCTCGACCAAGTGGCGCTGGCGGCGCCCACTTCGGCTTCGCGCGACAGCGCCGGC	156		

Db 61 CAGCCGCTCCGACCAAGTCGCGTCCGCGCCCACTTCGCTCCCGCCGACGCCCGGC 120
QY 157 CTTCGATGCGCAGCGCGCCGACCAAGTCGAGCTCCGAGATCGTCGAGAGCTCCACGAGAC 216
Db 121 CTTCGATGCGCAGCGCGCCGACCAAGTCGAGCTCCGAGATCGTCGAGAGCTCCACGAGAC 180
QY 217 CCCACCGACGAGCTCGTGAAGCTCAGCGGATACAGCTCACCCTCCGCTGAAGTTCGAGC 276
Db 181 CCCACCGACGAGCTCGTGAAGCTCAGCGGATACAGCTCACCCTCCGCTGAAGTTCGAGC 240
QY 277 GCGCGCCCGCAGAGGCGCGAGGCTCCGCGCTCCAGAACGACGACGAGATCCGCGACGCGTC 336
Db 241 GCGCGCCCGCAGAGGCGCGAGGCTCCGCGCTCCAGAACGACGACGAGATCCGCGACGCGTC 300
QY 337 GACAAAGAGCTGAGCTTCCTCAAGGCCGAGCTTCAGAACTCGGCTACGAGAGTCAACAC- 395
Db 301 GACAAAGAGCTGAGCTTCCTCAAGGCCGAGCTTCAGAACTCGGCTACGAGAGTCAACACG 360
QY 396 ----- 395
Db 361 GTGCGTCCGACGACGAGCGGCAAAATCTCGGATGCGCGACGCGTGAACGCTGACACTC 420
QY 396 -----GCGTTTCGATGCGCTCGCGCGACGACGAGGACTGAG 429
Db 421 GCTTGAACGCGCTGCCGCGCTTCGAGAGGATTCGATGCTCGCGACGACGAGGACTGAG 480
QY 430 GATGAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCCACG 489
Db 481 GATGAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCCACG 540
QY 490 TCCGCTCTGCTCTTAAAGCTGCGGAGCGCGCTCCAGAAACAGCTTCGCGCTGAAGTCTC 549
Db 541 TCCGCTCTGCTCTTAAAGCTGCGGAGCGCGCTCCAGAAACAGCTTCGCGCTGAAGTCTC 600
QY 550 CCGCGCGCATGATGATCGAGCTCAACTCGCTCAGCGTGCACCTGGCGCGTCCGCTC 609
Db 601 CCGCGCGCATGATGATCGAGCTCAACTCGCTCAGCGTGCACCTGGCGCGTCCGCTC 660
QY 610 GTGCTCTTGAAGCGCTCAACAACTTCTGAAACAACGCGATACGCCATCGTCCCTC 669
Db 661 GTGCTCTTGAAGCGCTCAACAACTTCTGAAACAACGCGATACGCCATCGTCCCTC 720
QY 670 CCGCGCTCCATTCGCGCGTCCGCGGACGCTCAGCGCGCTTCGATCAATCGCGCGGACATC 729
Db 721 CCGCGCTCCATTCGCGCGTCCGCGGACGCTCAGCGCGCTTCGATCAATCGCGCGGACATC 780
QY 730 ACCGATCAACCCGAGCTCAAGTTCAAGTTTGAACGAGGAAACGAGAAATCATGTT 789
Db 781 ACCGATCAACCCGAGCTCAAGTTTCAAGTTTGAACGAGGAAACGAGAAATCATGTT 840
QY 790 GCGCGGAGGCGATCTCGCTCTTTGGTCTCGAGGAGTC----- 828
Db 841 GCGCGGAGGCGATCTCGCTCTTTGGTCTCGAGGAGTCGCGATCGCGAGTCTGACT 900
QY 829 -----G 829
Db 901 GCAAGTGAAGTGTTCAGAGTCTCCAGTTTGTCTGACTGCCCTTTGTTCAATGCAATTGAC 960
QY 830 TCCCTCGCGCCGAGAGAGGCTCTCGATCTGATCAACGAAACGCGCGCTTCGCGCTCGATG 889
Db 961 TCCCTCGCGCCGAGAGAGGCTCTCGATCTGATCAACGAAACGCGCGCTTCGATG 1020
QY 890 CGACCTCATGCTGACGACTGACATGCTCTGCTCTCTCGAGGCGCTTCGAGGCTC 949
Db 1021 CGACCTCATGCTGACGACTGACATGCTCTGCTCTCTCGAGGCGCTTCGAGGCTC 1080
QY 950 TCACGATGAGGCGCATGCTGCGCGCAGAGGCGCTTCGCGCGCTTCATCCACGACGCTC 1009
Db 1081 TCACGATGAGGCGCATGCTGCGCGCAGAGGCGCTTCGCGCGCTTCATCCACGACGCTC 1140
QY 1010 GCGCGCCGACCCCGGCGCAGGTCGAGGTCGCGCGCAATCGCAGCTCTTTCGCGCT 1069

Db 1141 GCCCGCCGACCCCGCGCAGGTCGAGTCCGCGCAACATCCGACGCTCTTTCGCGCT 1200
QY 1070 CGTGGTTTGGCGTTGAGCAGAGAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 1129
Db 1201 CGTGGTTTGGCGTTGAGCAGAGAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 1260
QY 1130 GCCAGACCGGTCACCGCTCCGACGTCGCT----- 1161
Db 1261 GCCAGACCGGTCACCGCTCCGACGTCGCT-----CAGTTCGCGCGCGCTCG 1320
QY 1162 -----CAGTTCGCGCGCGCTCG 1180
Db 1321 CTCGCTCCGACCGCGCGCTCGAGACTTACGTTTTCGATTCAGTTCTCGCGCTCG 1380
QY 1181 TGAAGACATGATGACGCGCTTACTGACTCTCTGCTCGAGAACCAACACGACGACGAC 1240
Db 1381 TGAAGACATGATGACGCGCTTACTGACTCTCTGCTCGAGAACCAACACGACGACGAC 1440
QY 1241 ACCCGCTCTTGAAGTTCGAGAACCAACGAGACGCGCGCAATTCGAGGCGCTCG 1300
Db 1441 ACCCGCTCTTGAAGTTCGAGAACCAACGAGACGCGCGCAATTCGAGGCGCTCG 1500
QY 1301 CTGCTCATTTTCATGAGAGAGC----- 1325
Db 1501 CTGCTCATTTTCATGAGAGAGCAGGTCGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCT 1560
QY 1326 -----CAGGCTCGACTCGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1363
Db 1561 TGTGCTGATGTTCTTCTCTGAGGCTCGACTCGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1620
QY 1364 CGAGTGAACGAGTGTGTCAGGCTGCATGAACGCGCGCTCGCTCTGCTCTGCTCTGCTCTG 1423
Db 1621 CGAGTGAACGAGTGTGTCAGGCTGCATGAACGCGCGCTCGCTCTGCTCTGCTCTGCTCTG 1680
QY 1424 CCGAGACCGGTCGCTCAACTATCAGGCAAGGCTTGAACATTCAGATGCTGCTTACG 1483
Db 1681 CCGAGACCGGTCGCTCAACTATCAGGCAAGGCTTGAACATTCAGATGCTGCTTACG 1740
QY 1484 CTTGCG----- 1489
Db 1741 CTTGCGAGTGAAGCGCTGACGTTCTCGCGCGTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 1800
QY 1490 -----AGCTCGGCACTTTCGCAACCGCGCTGCTCTGCTCTGCTCTGCTCTGCTCTG 1530
Db 1801 TGACTTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
QY 1531 CCGCGAGATGAGCAACGAGCGCTGAACTCGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1590
Db 1861 CCGCGAGATGAGCAACGAGCGCTGAACTCGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1920
QY 1591 GCGGAGGCAACGAGCGCTCTCTC----- 1617
Db 1921 GCGGAGGCAACGAGCGCTCTCTCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 1980
QY 1618 -----CTTCTGCTCTGCACTGTAAGTCTGCAAGTCTGCAAGTCTGCAAGTCTGCAAGTCTG 1650
Db 1981 AGCGATGACTGCGCGATCCGAGGAGGTTCTGCTCTGCACTGTAAGTCTGCAAGTCTGCAAG 2040
QY 1651 GCGCTGACTTCCGCGCGATGAGTGAATTCAAGAGAGGTTGCAACCGCTTCTCCG 1710
Db 2041 GCGCTGACTTCCGCGCGATGAGTGAATTCAAGAGAGGTTGCAACCGCTTCTCCG 2100
QY 1711 ACTCTCTGCAAGCACTGCGGACTGAGCTGCAAGCACTGCGCTGCAAGTCTGCGCTGCAAGTCT 1770
Db 2101 ACTCTCTGCAAGCACTGCGGACTGAGCTGCAAGCACTGCGCTGCAAGTCTGCGCTGCAAGTCT 2160
QY 1771 AAGAGGCGCTCAACAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1830
Db 2161 AAGAGGCGCTCAACAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 1831 CACGACGCTTCTGTAAGCAACCGGAGCGCTGCTGAGAGCTCTCTGCTCTGCTCTGCTCTGCT 1890
Db 2221 CACGACGCTTCTGTAAGCAACCGGAGCGCTGCTGAGAGCTCTCTGCTCTGCTCTGCTCTGCT 2280

QY 1891 GCCACGTACACCTTACTGCGGTCAAACGGTGGAAAGTTGCTTCGGCCGAGAGGCCATC 1950
DB 2281 GCCAAGTACACCTTACTGCGGTCAAACGGTGGAAAGTTGCTTCGGCCGAGAGGCCATC 2340
QY 1951 TCGCTACGCGGAGGTGGCAACGGCTTCTGGGAGAGCGCGTCTTGGCAGGCGCGCGG 2010
DB 2341 TCGCTACGCGGAGGTGGCAACGGCTTCTGGGAGAGCGCGTCTTGGCAGGCGCGCGG 2400
QY 2011 CACGATACCTCTCGCCGCGCAACGGCTCTGTAAGTCTGTCGCGGAGAGCTCGG 2070
DB 2401 CACGATACCTCTCGCCGCGCAACGGCTCTGTAAGTCTGTCGCGGAGAGCTCGG 2460
QY 2071 GTGCAAGCGCGCGCGCGCAAGTCTTGTTCGCGGTGCAAGAGAGAGATCGGAGAAC 2130
DB 2461 GTGCAAGCGCGCGCGCGCAAGTCTTGTTCGCGGTGCAAGAGAGAGATCGGAGAAC 2520
QY 2131 GTCTCGCATCTACAGAGGCGCATCAAGAGCGGCGCATCAACAGTCTCTGTCAGATG 2190
DB 2521 GTCTCGCATCTACAGAGGCGCATCAAGAGCGGCGCATCAACAGTCTCTGTCAGATG 2580
QY 2191 CTGCGTAAAGCCCGGCAAGGCTCTGCTTAAGCGCGCGCTCACCCCAAGCAAGCTTTT 2250
DB 2581 CTGCGTAAAGCCCGGCAAGGCTCTGCTTAAGCGCGCGCTCACCCCAAGCAAGCTTTT 2640
QY 2251 CGACGTCTGTGTGTCGAAGAACGAGATTTCTCTCATACATGTGCGCTTACTCTCG 2310
DB 2641 CGACGTCTGTGTGTCGAAGAACGAGATTTCTCTCATACATGTGCGCTTACTCTCG 2700
QY 2311 CCGTCATCAAGTCTCTCAAGTCTTTTGTATCCCGCGTCTCT 2351
DB 2701 CCGTCATCAAGTCTCTCAAGTCTTTTGTATCCCGCGTCTCT 2741

RESULT 4
ARI99606 2475 bp DNA linear PAT 20-APR-2002
LOCUS ARI99606
DEFINITION Sequence 20 from patent US 6355468.
ACCESSION ARI99606
VERSION ARI99606.1 GI:20249680
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2475)
AUTHORS Yoshida,R.K. and Koobstra,A.B.
TITLE Phenylalanine ammonia lyase polypeptide and polynucleotide
JOURNAL Patent: US 6355468-A 20 12-MAR-2002;
FEATURES
source 1..2475
/organism="Unknown"
/mol_type="unasigned DNA"

ORIGIN
Query Match 67.3%; Score 1629; DB 6; Length 2475;
Best Local Similarity 78.0%; Pred.No.7.5e-189;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

QY 37 ATGGCCCTCTCTTGAATCGCTGCGCAACAGCTGCGCAACGGCTTACCAACGGCTCG 96
DB 1 ATGGCCCTCTCTTGAATCGCTGCGCAACAGCTGCGCAACGGCTTACCAACGGCTCG 60
QY 97 CAGCGCGCTCGACCAAGTGGCTGGCGGCCCACTTGGCTCTCCGCGCAGAC--CCG 154
DB 61 CAGCGCGCTCGACCAAGTGGCTGGCGGCCCACTTGGCTCTCCGCGCAGAC--CCG 120
QY 155 GCGTCGATGGCCAGCGCGCGCAACAGTGGCTGAGATGTGCAAGAGCTCTCAAG 214
DB 121 GCGTCGCTCTCCGCAACCAAGNNGAGAGCTCGACATGTMGAAGATCTCTGCGCG 180
QY 215 ACCCCACGAC---GACGTCTGAGCTCAGCGGGTACAGCTCAGCGTCCGTGACGTTG 271

DB 181 ACCCCACGNNACGNNNNNTGCAATCGACGGGTACACCTTCACTCGAGACTCG 240
QY 272 TCGGCGCGCCGGAAGGGGCGCAGGGTTCGCGTC-CAGAACGACGACGATCCGCGA 330
DB 241 TCGGCGCGCCGGAAGGGGCGCAGGGTTCGCGTC-CAGAACGACGATCCGCGA 300
QY 331 CGCGTCGACAAAGACGTGACTTCTCAAGGCCAGTTTAAAGTTCGTTCGCGAGTC 390
DB 301 AAGATGACAAAGNNTGAGTTCTCTCGNNCAGCTCAACAAAGNNTCTACGAGTC 360
QY 391 ACCAGGGGTTTGGTGGCTCGGCCGACACAGAGACTGAGATGACGCTCCAGAG 450
DB 361 ACGATGTTTGGCGGCTCGGCCGACACCGGACTGAGATGACGCTCCAGAG 420
QY 451 GCGCTCATGACACACAGCTTGTGCGCGTGAACCGACGTCGCTCTGCTTCAGGTC 510
DB 421 GCGCTCATGACACACAGCTTGTGCGCGTGAACCGACGTCGCTCTGCTTCAGGTC 480
QY 511 GGAAGCGGCTCGAGAACAGCTTTCGCTGAGGTGTCGCGGGCGCATGTCATCGCG 570
DB 481 GGAAGCGGCTCGAGAACAGCTTTCGCTGAGGTGTCGCGGGCGCATGTCATCGCG 540
QY 571 GTCAACTGCTCAGCGGTGACACTCGGCCGCTCGCTGCTCTTGAAGCGCTCAC 630
DB 541 GTCAACTGCTCAGCGGTGACACTCGGCCGCTCGCTGCTCTTGAAGCGCTCAC 600
QY 631 AACTTCTTGAACACCGCATACAGCCCATGTCCTCCCTCGCGGCTCTCATTCGCGCTG 690
DB 601 AACTTCTTGAACACCGCATACAGCCCATGTCCTCCCTCGCGGCTCTCATTCGCGCTG 660
QY 691 GCGGACCTCAGCCGCTCTGTACATGCGCGGCGCATCACGGTACACCGACGTCAG 750
DB 661 GCGGACCTCAGCCGCTCTGTACATGCGCGGCGCATCACGGTACACCGACGTCAG 720
QY 751 GTTACGTTTTCACAGAGGAAACGAGAGATATGTTTGGCGCGAGGCCATTCGCTC 810
DB 721 GTTACGTTTTCACAGAGGAAACGAGAGATATGTTTGGCGCGAGGCCATTCGCTC 780
QY 811 TTTGGTCTCGAGGAGTGTCTCGCGCGCAAGAGGTGTGTCGTCGTCGTCGTCGTC 870
DB 781 TTTGGTCTCGAGGAGTGTCTCGCGCGCAAGAGGTGTGTCGTCGTCGTCGTCGTC 840
QY 871 GCGGCTCTCGGCTCGATGAGGACCTTCACTGTCACAGCTCGCACATGCTCTGCTCTC 930
DB 841 GCGGCTCTCGGCTCGATGAGGACCTTCACTGTCACAGCTCGCACATGCTCTGCTCTC 900
QY 931 TCGCAGGCTTGAAGGCTCTCAAGGTGAGAGCCATGTCGCGCAGAGGGCTTCGCGG 990
DB 901 TCGCAGGCTTGAAGGCTCTCAAGGTGAGAGCCATGTCGCGCAGAGGGCTTCGCGG 960
QY 991 CCGTTCATTCACAGAGCTCTCGCGCGCACCGCGCAGGTCGAGGTCCGCGCAATC 1050
DB 961 CCGTTCATTCACAGAGCTCTCGCGCGCACCGCGCAGGTCGAGGTCCGCGCAATC 1020
QY 1051 CGACGCTCTCTTCGCGCTGTGCTGTTTGGCTTGAAGAGAGAGGTCAGAGTCAG 1110
DB 1021 CGACGCTCTCTTCGCGCTGTGCTGTTTGGCTTGAAGAGAGAGGTCAGAGTCAG 1080
QY 1111 GACGACGAGGACATTTCTTGCAGAGACCGGTACCGGCTCGACGTCGCTCAGTTCCTC 1170
DB 1081 GACGACGAGGACATTTCTTGCAGAGACCGGTACCGGCTCGACGTCGCTCAGTTCCTC 1140
QY 1171 GCGCCGCTCTGTAAGACATGATGACAGCTTCTGCTGCTGAGAAC--AAC 1227
DB 1141 GCGCCGCTCTGTAAGACATGATGATGACAGCTTCTGCTGCTGAGAAC--AAC 1200
QY 1228 ACGACGACGACACACCGGCTCTCGACGTCGAGAAACACACAGACGCGGACGCGGCAAC 1287
DB 1201 TCGACGACGACACACCGGCTCTCGACGTCGAGAAACACACAGACGCGGACGCGGCAAC 1260
QY 1288 TTTCAAGCGTGTGCTGTCTGATTTTCATGAGAGAACACAGGCTCGACCTCGCTCATC 1347
DB 1261 TTTCAAGCGTGTGCTGTCTGATTTTCATGAGAGAACACAGGCTCGACCTCGCTCATC 1320

QY 631 AACTTTGACACACCGCATACAGCCCATCGTCCCCCTCGGGCTCATCTGGGGCTCG 630
DB 601 AACTTCTCAACACGCGCATACCCCATCGTCCCCCTCGGGGACCATCTCGGGCTCG 660
QY 691 GGGGACCTGAGCCCGCTCTGTACATCGCCGGCCCATCACCGATCAACCCGAGCTCAG 750
DB 661 GGGGACCTCTCCCNCTCTMTACATCGCCGGCCCATCACCGATCAACCCGAGCANNAG 720
QY 751 GTTACGTTTGAAGAGGAAACGAGAAAGATCATGTTTGGCGCGAGGCCATTCGCTC 810
DB 721 GTTACGTTTGAAGAGGAAACGAGAAAGATCATGTTTGGCGCGAGGCCATTCGCTC 780
QY 811 TTGTGCTCGAGGAGTCTGCTCGCCCGAGAGAGGCTCGGCTGTGCTCAACGAGACG 870
DB 781 TTGTGCTCGAGGAGTCTGCTCGCCCGAGAGAGGCTCGGCTGTGCTCAACGAGACG 840
QY 871 GCGGCTCTCGGCTCGATGAGGACCTTCACTGCTGACAGCTGCTGCTGCTGCTC 930
DB 841 GCGGCTCTCGGCTCGATGAGGACCTTCACTGCTGACAGCTGCTGCTGCTGCTC 900
QY 931 TCGAGGCTTGAAGGCTCTCAAGGTGAGGAGCATGCTGAGGAGGAGGCTGCTGAG 990
DB 901 TCGAGGCTTGAAGGCTCTCAAGGTGAGGAGCATGCTGAGGAGGAGGCTGCTGAG 960
QY 991 CGGTTATCAACGAGCTGCTGCGCCGAGACCCCGGAGGTGAGGAGTGGCGGCAATC 1050
DB 961 CGGTTATCAACGAGCTGCTGCGCCGAGACCCCGGAGGTGAGGAGTGGCGGCAATC 1020
QY 1051 CGGAGGCTCTTCCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
DB 1021 CGGAGGCTCTTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 1111 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1170
DB 1081 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
QY 1171 GGGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1227
DB 1141 GGGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
QY 1228 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1287
DB 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
QY 1288 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1347
DB 1261 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
QY 1348 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1407
DB 1321 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
QY 1408 CCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1467
DB 1381 CCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
QY 1468 CAGCATCGGCTTACGCTTGGAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1527
DB 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
QY 1528 CAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1587
DB 1501 CAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
QY 1588 ACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1647
DB 1561 ACGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
QY 1648 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1707
DB 1621 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
QY 1708 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1755

DB 1681 NCNNGCTCTCTCAACGAGCACTTTGGGCGGCTTGAAGGAGGAGGAGGAGGAGGAG 1740
QY 1756 CTTCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1815
DB 1741 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
QY 1816 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1875
DB 1801 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
QY 1876 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1933
DB 1861 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
QY 1934 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1993
DB 1921 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
QY 1994 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2053
DB 1981 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
QY 2054 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2113
DB 2041 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
QY 2114 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2173
DB 2101 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
QY 2174 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2232
DB 2161 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
QY 2233 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2292
DB 2221 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
QY 2293 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2352
DB 2281 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
QY 2353 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2411
DB 2341 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
QY 2412 AAAAAA 2419
DB 2401 NNNNNN 2408

RESULT 6
E01759
LOCUS E01759 2151 bp RNA linear PAT 29-SEP-1997
DEFINITION DNA encoding L-phenylalanine ammonia-lyase.
ACCESSION E01759
VERSION E01759.1 GI:2170012
KEYWORDS JP 1988291583-A/1.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
REFERENCE 1 (bases 1 to 2151)
Fukuhara,N., Yoshino,S., Yamamoto,K., Nakajima,Y., Iwata,T.,
Matsumoto,N., Suzuki,M., and Makiguchi,N.
AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS
STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT
PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME
Patent: JP 1988291583-A 1 29-NOV-1988;
MITSUBI TOYOTSU CHEM INC
OS Rhodospiridium toruloides
COMMENT JP 1988291583-A/1

```

PD 29-NOV-1988
PF 22-MAY-1987 JP 1987123950
PI FUKUHARA NOBUHIRO, YOSHINO SADAO, YAMAMOTO KAORU, PI
PI NAKAJIMA YOSHIYUKI,
PI IWATA TOMOYUKI, WATANABE MIDORI, SUZUKI MAKI, PI MAKIGUCHI
NOBUOSHI
PC C12N9/88, C12N1/16, C12N1/18, C12N1/20, C12N15/00, C12P13/22, PC
(C12P13/22,
PC C12R1:19) , (C12P13/22, C12R1:865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=IRO 559;
CC *source: clone=PSM101;
FH key Location/Qualifiers
FH CDS 1..2151
FH FT /product='L-phenylalanine ammonia-lyase' FT
mat_peptide 1..2148
FT Location/Qualifiers
1..2151
/organism='Rhodospiridium toruloides'
/mol_type='genomic RNA'
/db_xref='taxon:5286'

ORIGIN
Query Match 49.5%, Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 165; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

FEATURES
source
Location/Qualifiers
1..2151
/organism='Rhodospiridium toruloides'
/mol_type='genomic RNA'
/db_xref='taxon:5286'

2Y 99 CGCGCTCGACCAAGTGGCTGGGGCCCACTTCGCTCCGCGGACAGCCGCGCT 158
DB 45 CTTCCATCCGCAAGCAGGCTGTCAATGCGCTTCAACCACTCGCACTCGCAGGCTC 104
2Y 159 CGATGCCACCGCGGCAACCACTGCACTGAGATGTGAGAGAGCTCTCTCAAGCC 218
DB 105 GCACTGCCCAACCAACCAAGTCAAGAGGTGCAATCGTCAAGAGATGCTCCCGCGCC 164
2Y 219 CACCGACACGCTGCTGAGCTCAAGGCTAAGGCTCAAGCTCCGTACCTTGTGGCGC 218
DB 165 GACCACTCGACGCTGCACTGCAAGGCTACTCGCTCAACCTCGGAGACGCTGCTCGGC 224
2Y 279 CGCCCGAAGGGGCGCAGGCTCCGCTCCAGACGACGACGACGACGACGCTCGA 338
DB 225 CGCGAGAAAGGCGAGGCTGTCCGCTCAAGAGACGAGACGAGATCGCTCAAGATGA 284
2Y 339 CAAGAGCGTGACTTCTCAAGGCCAGCTTCAAGACTCGTCTTACGAGTCAACAGG 398
DB 285 CAATCGGTCGAGTCTTCCGCTCGCACTCTCATAGCGCTTACGCGCTCAAGACTGG 344
2Y 399 TTTGGTGTGCTGGCGGACAGAGACTGAGAGTCAAGCTCAAGGCGCTCAT 458
DB 345 ATTTGGCGATCGAGAACCCGACCGGAGACGCAATCTCGCTTCAAGAGGCTTCTCT 404
2Y 459 CGAGACCAAGCTTCCGCGGCTGACCGCGCTCCGCTCGCTCTTCAAGCTCGAGCGG 518
DB 405 CGAGACCAAGCTTCCGCGGCTTCTCCGCTCGCTTCAAGCTCGCTTCCGCGCGCGG 464
2Y 519 CTTGAGAACAGCTTCCGCTTGAAGTGTCCGCGGCGCAATGTCTATCCGCTCAACTC 578
DB 465 TCTGAGAACTGCTTCCGCTTGAAGTGTCCGCGGCGCAATGTCTATCCGCTCAAG 524
2Y 579 GCTCAGCGTGGCACTCGCGCGCTCGCGCTCATGCTCTTGAAGCGCTCAACACTCTT 638
DB 525 CTTGAGCGCGGCACTCGCGCTTGAAGTGTCCGCTTGTCTTCAAGGCGCTCAACACTCTT 584
2Y 639 GAACCAAGCATACGCGCATGCTCCCTTCCGCGCTCATCTTCCGCTGCGGCGAGCT 658
DB 585 CAACCAAGCATACGCGCATGCTCCCTTCCGCGCTCATCTTCCGCTGCGGCGAGCT 644
2Y 699 CAGCCGCTCTGTACATCGCGGCGCATGACCGGTCACCCGAGCATCAAGGTTCAGT 758

```

```

DB 645 CTTCTCTCTCTCTCAATTCGACGCGCATACGCTTCAACCGGACAGCAAGGTGACGT 704
2Y 759 TTTGACAGAGGAAACCGAAGATCATGTTTGGCGCGAGGCATCTCGCTTTGGTCT 818
DB 705 CGTCCACGAGGCGAAGGAAAGATCTGTACGCCCGAGGCGCATGGCCCTTCAACCT 764
2Y 819 CGAGGAGTGTCTCTCGCGCGGAGAGGAGTCTGAGTGTGTAAGGAAAGGCGGCTCTC 878
DB 765 CGAGCGGCTGTCTCTCGCGCGGAGAGGAGTCTGAGTGTGTAAGGAAAGTCTCTC 824
2Y 879 CGCTCATGAGCGAAGCTTCAAGTGTGACAGACTGCAATGCTCTGCTCTCGAGAGC 938
DB 825 AGCATGATGAGCGAAGCTTCTGCTGACAGCGACAGCATGCTCTGCTCTCGAGTTC 884
2Y 939 CTTACAGGCTTCAAGTGTGAGGCGATGCTGCGGCGACAGGAGTCTGCTGCGGCTCAT 998
DB 885 GCTACAGCGCATGACGCTTCAAGGAGTGTGCGGCGACGCGGCTGCTTCAACCTCTCT 944
2Y 999 CCACGAGCTGCGCGCGGCGGACCGCGGCGAGTGTGAGGCTGCGGCGGACATCGCAGCT 1058
DB 945 TCAAGAGCTACGCGGCGCTTCAACCGAGAGTCAAGTGTGCGGAGAAATTCGCAAGCT 1004
2Y 1059 CTTTCCGCTGTGCTTGTGCTTGTGAGCAAGAGAGAGTCAAGTCAAGAGCAAGCA 1118
DB 1005 CTTGAGAGGAAAGCGCTTGTGCTGCTGCAACATGAGAGAGTCAAGTCAAGAGCAAGCA 1064
2Y 1119 GGGATTTCTTCCGAGGACCGGCTACCGGCTCCGACGCTGAGTCTGCTGCGGCGGCT 1178
DB 1065 GGGATTTCTTCCGAGGACCGGCTACCGGCTCCGACGCTGAGTCTGCTGCGGCGGCT 1124
2Y 1179 CGTGAAGACATGATGACGCTTCACTGACTCTGCTCTGAGAAACAA--CGAGAC 1235
DB 1125 CGTGAAGACATGATGACGCTTCACTGACTCTGACTGAGAGGCGGCGGCTGAGAGAC 1184
2Y 1236 CGACATCCGCTCTCTGACGCTTCAAGAAACAGAGACCGGCGGAGCACTTCCAGGCT 1295
DB 1185 CGACATCCGCTCTCTGACGCTTCAAGAAACAGAGACTTGCACACAGGCGGCGGCTTCAAGGCT 1244
2Y 1296 GTGAGCTGTCTGATTTGATGAGAGAACAGAGCTGACCTGCGCTTCAAGAGAGCT 1355
DB 1245 TCGGCTGTGCTCAACACATGAGAGAACTCGCTGAGGCTGCGGCGGAGAGCT 1304
2Y 1356 CAATTCACGAGTGAACCGAGTGTCAACGCTGCAATGACCGGCGGCTGCTTGTG 1415
DB 1305 CAATTCACGAGTGAACCGAGTGTCAACGCTGCAATGACCGGCGGCTGCTTGTG 1364
2Y 1426 CTTGCTGCGGAGAGACCGGCTGCTCACTATCAAGGAGGCTTGAATTCATTCG 1475
DB 1365 CTTGCTGCGGAGAGACCGGCTGCTCTTCACTATCAAGGAGGCTTGAATTCATTCG 1424
2Y 1476 TGCTTACGCTTGAAGCTGCGGCACTTTCGCAACCGGCTCACTTCTGTCAGGCGCG 1535
DB 1425 GAGCTACACCTGAGAGTGTGAGAACCTGCGGCACTGTCAGAGATGTCAGCGGCG 1484
2Y 1536 AGAGATGAGCAACAGGCGCTCACTGCTGCTTATCTCCGCGGCGGCACTGCGCA 1595
DB 1485 TGAATGCGCAACAGGCGGCTCACTGCTGCTTATCTCCGCGGCGGCACTGCGCA 1544
2Y 1596 GGGCAAGAGCTCTTCTCTCTCTGCTTCTGAGCTGCACTGCTTACTGCAAGCTTCCAGGCGCT 1655
DB 1545 GTCCAGAGAGCTCTTCTCTCTCTGCTTCTGAGCTGCACTGCTTACTGCAAGCTTCCAGGCGCT 1604
2Y 1656 CGACCTCGCGGAGTGAAGCTGCACTTCAAGAGAGTGTGAGCCGCTTCCGAGCTT 1715
DB 1605 CGACTTGGCGGAGTGAAGTTCAGATTCAAGAGAGTGTGAGCCGCTTCCGAGCTT 1664
2Y 1716 CTTGAGAGAGACTGCGCACTGCGCTGAGCTCA-----AGCAGCTTGGCT 1763
DB 1665 CATGACACGAGACTTGTGCTGCGCATGACCGGCTGCAACCTGCGGAGCAAGCTTGTGCA 1724
2Y 1764 CGAGGTCAAGAGGCGCTCAACAGGCTTTCAGACAGAGCAAGTACGACTCGAGCC 1823

```

Db 1725 GAAGGTGAACAAGAGCTGCGCAAGCGCTCGAGCAGACCACTGTAAGCACTGTCCTCC 1784
QY 1824 GCGCTGGACAGACGCTTCTGTAAGCAGACCGGACCGTGGTGGAGCTCTGTCCTC 1883
Db 1785 GCGCTGGACAGACGCTTCTGTAAGCAGACCGGACCGTGGTGGAGCTCTGTCCTC 1844
QY 1884 GCGCTTGGCAACGTCACCTTACTGCGGTCAACCGGTGAAAGTGGTGGCGGAGAA 1943
Db 1845 GTC-----GCTCTGCTGCGCGCGGTCAAGCGCTGAAAGGTGCGCGCGCGAGTC 1895
QY 1944 GCGCATCTGCTCAGCGCGGAGGTGCGAACCGCTTCTGAGCAGCGCGCTTTCGACAGC 2003
Db 1896 GCGCATCTGCTCAGCGCGGAGGTGCGAACCGCTTCTGAGCAGCGCGCTTTCGAGC 1955
QY 2004 GCGCGGACGACGATCTCTGCGCGGACGCGGCTCTGACTGCTTCTGCGCGGAGGA 2063
Db 1956 GCGCGGCTCTGCTGCTCTGCTGCGCGGACGCTCAAGTCTCTGACGCTTCTGCGGAGGA 2015
QY 2064 GCTCGGCTGACAGGCGCGCGCGGACGCTGTTGTGGCGGTGACAGAGACGATCGG 2123
Db 2016 GCTTGGCGTCAAGGCGCGCGCGGACGCTTCTGCGGACAGAGAGGTGACGATCGG 2075
QY 2124 GAGCAAGCTCTGCGGACGCTTACAGAGGACCATGACAGAGCGGACGATCAACGACGCTCT 2183
Db 2076 CTCGAAGCTCTCAAGATCTTACAGAGGACCATGACGAGTGGGACGATCAACGATCTCTCT 2135
QY 2184 CAAGATGCTCGGGA 2198
Db 2136 CAAGATGCTCGCTTA 2150

RESULT 7
AR205032 2151 bp DNA linear PAR 20-JUN-2002
LOCUS AR205032
DEFINITION Sequence 9 from patent US 6368837.
ACCESSION AR205032
VERSION AR205032.1 GI:21502510
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2151)
AUTHORS Gatenby,A.A., Saitiiani,S., Tang,X.-S., Qi,W.Wei. and Vannelli,I.T.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent: US 6368837-A 9-09-Apr-2002;
FEATURES
Location/Qualifiers
1..2151
/organism="Unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 165; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 39 GCGCGCTCGAACGATGCGCTGCGGCGGACCTTTCGCTCTCGCGGACGCGCGGCT 158
Db 45 CCTCGCATCGCAAGAGAGGCTGCAATGAGGCTTCGACCAACTGCAAGTGGCAGGCTC 104
QY 159 CGATGGCAGCGCGGACGACAGTGCAGCTGAGATGTGAGAGAGTCTCTAGCAGCC 218
Db 105 GCACCTGCGCAACCCAGGTCAAGCAGAGTGCAGATGTGAGAAAGTGTCCCGCGGC 164
QY 219 CACGAGCAGCGTGTGAGGTCAGCGGTACAGCTCAGCTCCGTCGAGAGTGTGGCGGC 278
Db 165 GACGACCTCGACGCTGACCTGACGAGGCTACCTGCTCAGAGAGTGTCTTCGCGC 224
QY 279 GCGCGCAAGGCGGACGAGTTCGCGTTCAGAAAGCAGACGAGATTCGCGCAGCGTCGA 338
Db 225 GCGCGAAGGCGGAGGCTGTCGCGTCAAGACAGAGGATCGCGTCAAAAGATTGA 284
QY 339 CAAGAGCGTCACTTCTCAAGGCGCGGCTTCAAGAGTGTGAGAGTCAACCGGCG 398

Db 285 CAATTCGTCGAGTCTTTCGCTCGCAACTCTCCATGAGGTCTACGCGTCAAGACTCG 344
QY 399 TTTGGGTGCTCGGCGCGACACAGAGACTGAGATGAGTCAAGCTCCAGAAAGCGCTCAT 458
Db 345 ATTGGCGGATCCCAACACCGCACCGAGAGCGCATCTCGCTCAAGAGGCTTCTCT 404
QY 459 CGAGACGAGCTTTCGCGGCGTACGCGGAGTCCGTCGCTTTCAGGCTGCGAGCGG 518
Db 405 CGAGACGAGCTTTCGCGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 464
QY 519 CTTGAGAAACGCTTTCGCTGAGAGTTCGCGGCGGCGGATGCTATCCGCTCACTC 578
Db 465 TCTGAGAACTGCTTTCGCTGAGGTTGTCGCGGCGGATGCTATCCGCTCAACG 524
QY 579 GCTACGAGTGGGCACTCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
Db 525 CTTGACCGGCGGACCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
QY 639 GAACCCACGATCAAGCGGCTATGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
Db 585 CAACACGAGATCACCGGCTATGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
QY 699 CAGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758
Db 645 CT 704
QY 759 TTTGACAGAGGAAACCGAAGATCAATGTTGCGCGCGAGGCTATCTGCTTGTGCT 818
Db 705 GGTCAAGAGGAGGAGAGAGATCTGTAAGGCTGTAAGGCTGTAAGGCTGTAAGGCT 764
QY 819 CGAGGAGTGTCTCTGCGGCGGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878
Db 765 GAGCGCGCTGCTCTGCGGCGGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
QY 879 GCGCTGATGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 938
Db 825 AGCATGATGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
QY 939 CTTGACGCTCTGACGCTGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 998
Db 885 GCTCAGCGGATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 999 CCAAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1058
Db 945 TCAGAGCTACGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
QY 1059 CTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
Db 1005 CTTGAGGAGAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
QY 1119 GGGATCTCTGCGGAGGACCGGCTACCGGCTGCGGAGTGTGCTGCTGCTGCTGCTGCT 1178
Db 1065 GGGATCTCTGCGGAGGACCGGCTACCGGCTGCGGAGTGTGCTGCTGCTGCTGCTGCT 1124
QY 1179 CGTGAAGAGATGATGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1125 GGTGAAGAGATGATGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
QY 1236 CGACACCGCTCTGAGGTCGAGAAAGAGACCGGCGGCGGCACTTCAAGGC 1295
Db 1185 CGACACCGCTCTGAGGTCGAGAAAGAGACCGGCGGCGGCACTTCAAGGC 1244
QY 1296 GTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
Db 1245 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
QY 1356 CAATTCAGCAGTGCAGCGAGTGTCTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
Db 1305 CAATTCAGCAGTGCAGCGAGTGTCTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
QY 1416 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1475
Db 1365 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424

Oy	1111	GGGCACTTCTTCGCGAGAGACCGCTACCCGGCTCCGACAGCTGCCTCAGTCTCCGACCGCT	1178
Oy	1065	GGGCACTTCTCCGCGAGAGACCGCTTACCCTTTCGCAAGCTCTCTCAGTGGCTGGCCGCT	1124
Oy	1179	CGTGGAGACATGATGACAGCGCTTACTGACTCTCTCGCTCGAGACAAACA---CGAGAC	1235
Db	1125	CGTCAACGACCTCATTTCAAGGCCACGCGCTCTCACCATGAGAGCGCGCAGTCAACAC	1184
Oy	1236	CGACAAACCCGCTCTCTCGAGCTCGAGAACAAACGACACCGCGACGCGGCGGCACTCCAGC	1295
Db	1185	CGACAAACCTTCTCATCGAGCTCGAGAACAAACTTTCGACACACGCGGCGCAATTTCCAGGC	1244
Oy	1296	GTCGCTGCTTCGATTTTCGATGAGAAAGACCGACTCGACCTCGCCTCATCGGACGT	1355
Db	1245	TGCGGCTGTGGGCAACACCATGAGAGACTCGCTCGGGCTCGCCCAAGTCGGAGCT	1304
Oy	1356	CAACTTCATGACGTGACCGAGTTGCTCAAGCTGCGCATTAACCGGCGGCTGCTTCGT	1415
Db	1305	CAACTTCAGGCACTTCACCGAGATCTCAAGCCGCGCATTAACCGGCGGCTCTCCCTCTG	1364
Oy	1416	CTCGCTGCGGAGACCGGTGCTCACTATCAGCGCAAGGCTTGAATTCATCAATGAC	1475
Db	1365	CTCGCGGCGGAAAGACCCCTCGCTCTCTTACACTGCAAGGCTCGACATGCGCGTGC	1424
Oy	1476	TGCTTAAGTTTGGAGAGCTCGGCGACCTTGGCAACCGCGGTACATCTTGTCACGCGC	1535
Db	1425	GCGGTACACTTCGGAGTTGGAGACTTCGCAACTTCGTBACACGCAATGTCCAGCCGC	1484
Oy	1536	AGAGATGGGCAACCAAGCCGTCAACTCGCTCGCTCTCATCTCGCGCGCGCACTGCGA	1595
Db	1485	TGAGATGGGCAACCAAGCGGTCAACTCGCTTCGCTCATCTCGGCTCTGTCGACACACGA	1544
Oy	1596	GCGCAACGACAGTCCCTTTCATCTCTTTCGCTCGACCTGCACTGTATCGACGCTCCAGCGCT	1655
Db	1545	GTCCAACGACAGTCCCTTCTCTCTCTCTCTCGCACCCCACTTACGTGGCTTCTCCAGCAAT	1604
Oy	1656	CGACTCGCGCGATGAGAGCTTCGACTTCAAGAAAGAGTTGACCCGCTCTTCCGACTCT	1715
Db	1605	CGACTTCGCGCGACCGAGTTCGAGTTCAAGAAAGAGTTGCGCCACGCACTGCTTCCT	1664
Oy	1716	CCGCCAGCGACCTCGGCACTGGGCTCGAAGTCA-----ACGCACTTGGCCT	1763
Db	1665	CATCGACACGACTTTGGCTTCGCAATGACCGGCTCGAAGCTTCGCGAGAGACTGTGCGA	1724
Oy	1764	CGAGTCAAGAAAGCGCTTCAAAAGCTCTCGAGACAGACGACGTACACTTGAAGCC	1823
Db	1725	GAAAGTGAACAAAGAGCTTCGCAAGCGCTCGACACCAACCACTTCGACACTGTGCC	1784
Oy	1824	GCGCTCGACAGAGCGCTTCTGTTAGCGAGACCGGACAGTGTGAAGCTCTCTCGTCTC	1883
Db	1785	GCGCTCGACAGAGCGCTTCTCTTTCGCGCGCGACCGCTGTGAGAGTCTCTTCGTGAC	1844
Oy	1884	GCCCTTCGCAACGTACCCCTTACTGCGCGTCAACGCGTGAAGGTTGCTCGGCCGAGAA	1943
Db	1845	GTC-----GCTCTTGCTCGCGCGCTGTCACAGCTTGAGAGTGC CGCGCGCAATC	1895
Oy	1944	GCGCATCTGCTCAACGCGAGAGTGGCAACCGTTTGGCAACACGCGCTTTCGACAGC	2003
Db	1896	GCGCATCTGCTCAACCGCGCAAGTCCGAGAGACTTCTGTCGCGCGGTGACACTGTC	1955
Oy	2004	GCGGAGGACGCACTCTCTTCGCGCGGACGCGGCTCTGTACTGTTCTGTGCGGAGGA	2063
Db	1956	GCGCGGCTCTGTACTCTCTCGCGCGCACTTCAGATCTCTTACGCTTGTGTCGGAAGA	2015
Oy	2064	GCTCGGCGTGAAGCGCGCGCGCGAGAGTGTTCGCGGTGACAGCAAGACATCGG	2123
Db	2016	GCTTGGCGTCAAGCGCGCGCGGAGAGCTTCTCTTCGGAAGCAAGAGTGAACATCGG	2075
Oy	2124	GAGCAACGTCTTGCGCACTTTCAGAGCGATCAAGACGCGCGCATTAACACAGTCTTCT	2183
Db	2076	CTCGAAGCTTCAAGACTTTCAGAGCGATCAAGTCAAGTTCGCGGAGGATCAACAGAGTCTCT	2135

[illegible]

ORIGIN

Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CCGCGCTCCGACCACTCGCTCGCGGCCCACTTCGCTCTCCGCGACGCCGCT 158
DB 45 CGTCGATCCGGAAGCAGCTGTCAATGGCGCTTCACCAACCTCGAGTCGACGCTC 104
QY 159 CGATGACCAAGCGCGCCACCACTCCAGTCGAGATGTCGAGAGCTCTCAAGCA 218
DB 105 GCACCTGCCCAACCAACCAAGTCACGACGATCGATCGAGAAAGTCTCGCGGCC 164
QY 219 CACCGACGAGTCGTCGAGTCAGCGGGTACAGCTCCAGCTCGTGAAGTTCGCGCG 278
DB 165 GACCGACTCGACGCTCGAATCGAGGCTACCTCGCTCAACTCGGAACGTCGTGCGC 224
QY 279 CGCGCCGAAGGGCGCAGGGTCCGCTCGACAGCAGAGATCCGCGCAGCGCTCGA 338
DB 225 CGGAGGAAAGGGAGGCTGTCGCGGTCAAGGACAGCAGAGATCCGCTCAAGATTGA 284
QY 339 CAAGAGCTGCACTTCCTCAAGGCCAGCTTCAGAACTCGGTCACGAGTCAACAGGG 398
DB 285 CAATGCTGAGGTTCTTGCTTCGCTCGCAACTCTCCATGAGGCTTACGCGCTCAAGCTGG 344
QY 399 TTTCGGTGGCTCGCGCGACAGAGACTGAGATGCACTCAGCTTCAGAAAGCGCTCAT 458
DB 345 ATTGGCGATCCGAGACACCCGACAGAGAGCCCATCTCGCTCCAGAAAGGCTCTCT 404
QY 459 CGAGCACAGCTCTCGCGGCTGACCCGCAAGTCCTGCTCTTCAAGCTTCGACGCG 518
DB 405 CGAGCACAGCTCTCGCGGCTGACCCGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 464
QY 519 CTTGAGAACAGCTTCGCTCGAGGTGTCGCGCGCGCATGTCATTCGCGCTCACTC 578
DB 465 TCTCGAAGCTGCTCTTCCCTCGAGGTTTTCGCGCGCATGCAATTCGCGCTCAAG 524
QY 579 GCTCAGCGTGGCACTCGCGCGCTCGCTGCTGCTCTTGAAGCGCTCAACATCTT 638
DB 525 CTGTGACCCGCGCGCACTCGCGCTGCTCGCTGCTCTGCTCTGCTCTGCTCTGCT 584
QY 639 GAACACCGCATCAAGCCCATGTCCTCCCTCGCGCGCTCACTTCGCGCTCGCGCATCT 698
DB 585 CAACACCGCATCAAGCCCATGTCCTCCCTCGCGCGCATCTTCGCTGCGCGCATCT 644
QY 699 CAGCCGCTCTGTCATTCGCGCGCATCAAGCTCAACCCGACGTCATGAGTTTCACT 758
DB 645 CTCTCTCTCTCTCAATTGAGCGCGCATCAAGGTCACCGAAGAGTCACT 704
QY 759 TTTGACAGGAGAACCGGAAAGATATGTTTCCGCGGAGGCACTTCGCTCTTTGCT 818
DB 705 CGTTCACAGAGGAGAAAGAAATCTGTAGCCCGGAGGCGATGCGCTCTTCAACCT 764
QY 819 CGAGCAGTCTGCTCGCGCGCAAGAGGTCGCTGCTGCTGTCAGGAAAGCGCGCTC 878
DB 765 CGAGCCGCTGCTCTCGCGCGCAAGAGGTCGCTGCTGTCAGGAAAGCGCGCTC 824
QY 879 CGCTCGATGCGACCTTCACTGTCAGACTCGACATGCTCTGCTCTCTCGAGG 938
DB 825 ACGATTCGATGCGACCTTCTGCTGTCAGCAGCATATGCTCTGCTCTCTCGAGTC 884
QY 939 CTTGAGCGCTTCAGGTCGAGGAGGTCAGTTCGCGCAGAGGCTCTGCGCGCTTCA 998
DB 885 GCTTCAGGCGATGAGGTCGAGGAGGTCGAGGTCGAGGCGCGCTCTTCAACCT 944
QY 999 CCACGAGCTTCGCGCGCGCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 1058
DB 945 TCAGACGTCAGCGCGCTCTCAACCGAGCAGATGAGGTCGAGGTCGAGGTCGAGGTC 1004
QY 1059 CTTTCGCGCTGCTGTTGCGCTTGAAGCAGAGGAGGTCAGGTCAGAGCAGAGC 1118
DB 1005 CTTGAGGAGGAGCGCTTGTCTTCACCATGAGGAGGTCAGGTCAGAGCAGAGC 1064

QY 1119 GGGCATCTTGCCAGGACCGCTACCGGCTCCGACGTCGCTCAAGTTCCTCGCGCGCT 1178
DB 1065 GGGCATCTTGCCAGGACCGCTACCGGCTCCGACGTCGCTCAAGTTCCTCGCGCGCT 1124
QY 1179 CGTGAGGACATGATGACGCTTACTGACTCTCTGCTCGAGAACAAACA---CGACGAC 1235
DB 1125 CGTCAGACCTCATTTACCGCCCAAGCGCTCTTACCATGAGGCGCGGACGTCAGCAG 1184
QY 1235 CGACACCCGCTCTCTGACGTCGAGAACAGCAGACCGCGCAGCGGCGGACCTTCCAGG 1295
DB 1185 CGACACCCCTTCTTCAAGCAGTCGAGAACAAAGACTTCGACACAGCGGCGGCAATTTCAAGG 1244
QY 1286 GTGCGCTGTCGATTCGATTCGAGAACACAGGCTGCACTCGCTCTCACTCGGCAAGCT 1355
DB 1245 TGCGCGCTGTCGACCAACACAGAGAGAGCTCGCGCTCGGCGCTCGCCAGTCAGT 1304
QY 1356 CACTTCACGAGTCACCGAGTTCGACAGCTTCGACAGCTTCGACAGCTTCGACAGCT 1415
DB 1305 CACTTCACGAGTCACCGAGTTCGACAGCTTCGACAGCTTCGACAGCTTCGACAGCT 1364
QY 1416 CCGCGCTCGAGAGCCGCTCGCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1475
DB 1365 CCGCGCGCGCAACCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
QY 1476 TGTTCAGCTTCGAGGCTCGGCGCACTTCGCAACCCGCTCACTCACTCTCTCTCTCT 1535
DB 1425 GAGCTACACCTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 1484
QY 1536 AGAGTGGGACCAAGGCGCTCACTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
DB 1485 TGAATGCGCAACCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 1544
QY 1596 GGGCAACGAGCT 1655
DB 1545 GTCCAGCAGCT 1604
QY 1656 CGACCTTCGCGCGATGAGGTCGATTCGAAAGAGGTCGAGGTCGAGGTCGAGGTCGAG 1715
DB 1605 CGACCTTCGCGCGAGCCGATTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 1664
QY 1716 CTTCCAGAGACCTTCGCGCACTGCGCTCGACGTCGCA-----AGCACTTCGCT 1763
DB 1655 CATGACAGACTTTGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1724
QY 1764 CGAGGTCAAGAGGCGCTCAACAGGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
DB 1725 GAGGTCAACAGAGGCTTCGCAAGGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1784
QY 1824 GCGCTGCGAGAGGCTTCGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1883
DB 1785 GCGCTGCGAGAGGCTTCGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
QY 1884 GCGCTTCGCAAGCTCACT 1943
DB 1845 GTC-----GCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 1895
QY 1944 GGGCATTCGCTCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2003
DB 1896 GGGCATTCGCTCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 1955
QY 2004 GCGCGCGAGAGATCTCTCGCGCGAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2063
DB 1956 GCGCGCGCTCTGATCTTCGCGCGAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2015
QY 2064 GCTCGCGCTCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 2123
DB 2016 GCTTCGCGCTCAAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 2075
QY 2124 GAGCAAGCTTCGCGCATTCAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 2183
DB 2076 CTGCAAGCTTCGCAAGATTCAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 2135

2y 2184 CAAGATGCTGGCGTA 2198
2b 2136 CAAGATGCTGCTTA 2150

RESULT 11
LOCUS RFPALGNA 2315 bp RNA linear PLN 23-MAR-1993
DEFINITION R.toruloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION X12702.1 GI:288350
VERSION X12702.1
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
1 (bases 1 to 2315)
REFERENCE 1
AUTHORS Unknown.
JOURNAL Unpublished
COMMENT (Patented) BP 0 260 919 A1.
FEATURES
location/Qualifiers
1..2315
/organism="Rhodospiridium toruloides"
/mol_type="mRNA"
/strain="IFO 559"
/db_xref="taxon:5286"
/clone="psw13"
1..2315
/gene="PAL"
1..2151
/gene="PAL"
/codon_start=1
/product="L-phenylalanine ammonia-lyase"
/protein_id="CA31209.1"
/db_xref="GI:288351"
/db_xref="GOA:P11544"
/db_xref="SWISS-PROT:P11544"
/translation="MAPSLDISHSFANGVAKAKQAVNGASINLAVGSHLPTQVTO
VDIVKMLAAPDSTLEIDGSLNLDGVNAAKRGKRVKDSDETRSKIDKSEVETLR
PLEVNGVYGVTFGGSGADTEDEAISLQALHEHQCVLPSPSPFRIGRLNLSI
SVLVRGAMTIRVNSLTRGSHAVRVLVLEALTNFNGITPIVPLRGTTISAGDLSPL
SYIAAITSGHPSKVHVHVEGKELIYAREMALFNLEPVLDGKXGLGVNGTAVSA
SMATLALHDAMLSLSOSLTVAMVEAMVHAGSFMFPDTPRPHPTGIVAGNRK
LLEGRFVHHEEVKVDDEGILKQRPYRTPSPMLGFLVSDLIHAHNVLTTEGQ
STTNPPLIDVENKTSHGNGNFQAAAVANTWEKTRLGIAQIKLNTQUTEMLVAGNR
GLPSCLAEDPSLSYHCKGLDIAAAVTSSEIHLAVTTHVPPAMANOAVNSLMI
SARTEISNDVLSLILATHLKCYQLDIPRALIEFEKQGPALIVSLIDQHFQSAMTG
SNRREIYKXNKLAKRLKQTNQYDLVPRHDAFPAAGTVAIVELSTSLIAVNA
WKVAASBAISLTROVRETFMSASTSPALSYLSPPTQILYAFVBEELGVKARQDV
PLGKEVITGSNVSKITIYALNSGRINNVLMKTLA"
2315
/gene="PAL"

polYA_site
/gene="PAL"

ORIGIN
Query Match 49.5%; Score 1196.6; DB 8; Length 2315;
Best Local Similarity 74.0%; Pred. No. 2.4e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
QY 99 CCGCGCTCCGACCAAGTCGCGTCCGCGCCCACTTCGCTCCGCGCAAGCGCGCT 158
Db 45 CCGCGATCCGCAAGGCGAGGCTGTCAATGCGCGCTCGACCACTCCGAGTCGAGGCTC 104
QY 159 CGATGGCCACCGCGCGACCACTCCGAGTGTGAGAGAGAGTCTCTCAGACCC 218
Db 105 GCACTGCCCAACCCAGTCACGCAAGTCGATCTGAGAGAGATGCTCGCGCGCC 164
QY 219 CACCGACGATGCTGAGACTCAGCGGTAACAGCTCCGCTCGGTGAGTTCGCGCG 278
Db 165 GACCTGACTGAGGCTGAACTGACGCTACTCGCTCAACTCGAGAGAGTCTCTCGGC 224
QY 279 CGCCCGCAAGGCGCGAGGTCGCGTCCAGACGACGACGAGATCCGCGCAGCGCTCGA 338
Db 225 CGGAGAGAGGCGCGCTGTCCGCTCAAGGACGACGACGAGATCCGCTCAAGATGGA 284

QY 339 CAAGAGCTGCACTTCCTCAAGGCCAGCTTGAAGACTCGATCTACGAGATGACACGAG 398
Db 285 CAATCGGTGAGATTCCTGCGCTCGCACTCTCCATGAGCGTCTAGGGGTACACGCTGG 344
QY 399 TTTCGTGCTCGCGCGACACGAGACTGAGATGCACTCAGCTTCAGAAAGCGCTCAT 458
Db 345 ATTGGCGGATCCGACAGACCCGACCGACGACGACATTCGCTCCAGAAAGCTTCCT 404
QY 459 CGAGCACCACTCGCGCGGCTGACGCGGACGTCGCTGCTGCTTCAAGGTGAGACGCG 518
Db 405 CGACACCACTCTGCGGCTTCTCCCTTGTGTCTGACTGCTTCCGCTCGCGCGCG 464
QY 519 CTTGAGAACACGCTTCGCTGAGGTGCTCGCGGCGCCATGTCATTCGCGTCACTC 578
Db 465 TCTGAGAACCTCGGTTCCCTTCGAGGTGTCGCGGCGCCATGACATCCGCGTCAACG 524
QY 579 GCTCAACGCTGCGCACCTCGCGCGCTCGCGCTCGCTGCTGAGGCGCTCACCAACTCT 638
Db 525 CTTGACCCGCGGACCTCGGCTGCTCGCGCTGCTGCTGAGGCGCTCACCAACTCT 584
QY 639 GAACCAACCGATCACGCGCCATCGTCCCTCCGCGGCTTCATCTCGCGTGGCGACCT 698
Db 585 CAACCAACCGATCACCCCATCGTCCCTCCGCGGACCATCTGCGTGGCGGACCT 644
QY 699 CAGCCGCTCTGTAATGCTCGCGGCGCCATCACCGTCAACCCGAGTGAAGTTACGT 758
Db 645 CTTCTCTCTCTCAATTGACGCGGCATCAGGTCACCCGGAAGAGTGCACGT 704
QY 759 TTGACAGAGGAACCGAAGATCATGTTGCGCGGAGGCGCATCTCGCTTTGCT 818
Db 705 CTTCAAGAGGCAAGAGATCTGTACGCGCGGAGGCGATGGCGCTCTTCAACT 764
QY 819 CGAGCGATGCTCTCGCGCGGCGGAGAGGCTCTGCTGCTGTCACGGAAGCGCGCTTC 878
Db 765 CGAGCGATGCTCTCGCGCGGCGGAGAGGCTCTGCTGCTGTCACGGAAGCGCGCTTC 824
QY 879 CGCTCGATGCGGACCTCAAGTCTGACGACCTGCGACATGCTGCTGCTCGAGCG 928
Db 825 AGCATGATGCGGACCTCGCTGTCGACGACGACATATCTTCTGCTCTCGAGTC 884
QY 939 CTTGACGCTCTCAAGTGTGAGGCGCATGTCGACGACGAGGCTGTCGCGCGCTTCAT 998
Db 885 GCTCAAGGCGATGACGCTGCAAGGATGTCGCGGACGCGGCTGCTTCAACCTTCCT 944
QY 999 CGAGAGCTCTGCGCGCGGACCGCGGCGAGTCAAGTGGCGCGGACATTCGCGAGCT 1058
Db 945 TCAGACCTCAAGGCGGCTCAAGGCGGAGTCAAGTGGCGGAGATTCGCGAGCT 1004
QY 1059 CTTTCGCGCTCGTCTGCTGCTTGAAGCAAGAGAGGTCAAGTCAAGAGCGACGA 1118
Db 1005 CTTGAGGAGAGCGGCTTGTGCTGCTCACTGAGAGAGAGTCAAGTCAAGAGCGACGA 1064
QY 1119 GGGCATTTTCGCGAGGACCGGTAACCGGCTCGGACGTCGCTCAAGTTCTCGCGCGCT 1178
Db 1065 GGGCATTTTCGCGAGGACCGGTAACCGGCTCGGACGTCGCTCAAGTTCTCGCGCGCT 1124
QY 1179 CGTGAAGACATGATGACGCGCTCTGAGTCTGCTGCTGAGAAACA---GACGAC 1235
Db 1125 CTTGAGACCTCAATTCACGCTCAAGGCTCTCAATGAGGCGGCGAGTGAAGAC 1184
QY 1236 CGACACCGGCTCTGAGCTGAGAAACAAGCAGACCGGCGGCAACTTCAGGC 1295
Db 1185 CGACACCGGCTCTGAGCTGAGAAACAAGCTTCAGACGAGGCGGCAATTCAGGC 1244
QY 1296 GTGCGCTGCTGATTTGATGAGAAAGACCAAGCTGCACTCGCGCTCATCGGCAAGCT 1355
Db 1245 TCCCGCTGAGCAACACATGAGAAAGACTCGCTGAGGCTGCGCAAGTCCGCAAGCT 1304
QY 1356 CAACCTCAAGAGTCAAGAGTTCGCAAGCTCAAGCTGAGTCAAGCTGAGCTGCTTCG 1415
Db 1305 CAACCTCAAGAGTCAAGAGTTCGCAAGCTCAAGCTGAGTCAAGCTGAGCTGCTTCG 1364

QY 1416 CTTGGTGGCGAGGAGACCCGTCGCTCACTATCAAGCGAGGGCTTGGACATTCAATCCG 1475
 DB 1365 CTTGGTGGCGAGGAGACCCGTCGCTCACTATCAAGCGAGGGCTTGGACATTCAATCCG 1424
 QY 1476 TGGTTACGCTTCGAGCTCGGCACTTGGCCAAACCGGTCACTACCTTCGTCAGCCCGC 1535
 DB 1425 GGGGTACACCTCGGAGTTGGACACCTGGCCAAACCTTGAAGAGCATGTCCAGCCGGC 1484
 QY 1536 AGGATGGGCAACAGGCGCTCACTCGCTGCTCTCATCTCCGCGCGCGACTGCGCA 1595
 DB 1485 TGGATGGGCAACAGGCGCTCACTCGCTGCTCTCATCTCCGCGCGCGACTGCGCA 1544
 QY 1596 GGGCAACGAGCT 1655
 DB 1545 GTCGACGAGCT 1604
 QY 1656 CGACCTCGCGCGAGTGAAGCTGACTTCAAGAGCATTTGCGCCAGCCATCTCTCTCT 1715
 DB 1605 CGACTTGGCGCGAGTGAAGTTCAGAGTCAAGAGCATTTGCGCCAGCCATCTCTCTCT 1664
 QY 1716 CTTCCAGCAGCACTCTCGGCACTGAGCTGAGCTCA-----ACGCACTTGGCT 1763
 DB 1665 CATCGACGAGCACTTGGCTCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1724
 QY 1764 CGAGTCAAGAGGCGCTCAACAGCGCTCTGAGCAGCAGCAGTACGACCTCGAGCC 1823
 DB 1725 GAAGTGAACAGAGCGCTGCGCAAGCGCTGAGCAGCAGCAGTACGACCTCGAGCC 1784
 QY 1824 GGGCTGGCAGAGCGCT 1883
 DB 1785 GGGCTGGCAGAGCGCT 1844
 QY 1884 GGCCTCTGCAAGCTCACT 1943
 DB 1845 GTC-----GCTCTGGCTGGCGCGCTGCAAGCGCTGAGAGTTCGCGCCGCGCGAGTTC 1895
 QY 1944 GGGCATTTGCTCAAGCGCGGAGGTCGCAACCGCTTCTGAGCAGCGCGCTTCTGAGAGC 2003
 DB 1896 GGCATTTGCTCAAGCGCGGAGGTCGCAACCGCTTCTGAGCAGCGCGCTTCTGAGAGC 1955
 QY 2004 GCGCGCAGCAGCATCT 2063
 DB 1956 GCGCGCAGCAGCATCT 2015
 QY 2064 GCTCGAGCTGCAAGCGCGCGCGCGAGTGTGTCGCGCTGAGCAGAGAGAGCATCGG 2123
 DB 2016 GCTCGAGCTGCAAGCGCGCGCGCGAGTGTGTCGCGCGAGAGAGAGTGTGAGATCGG 2075
 QY 2124 GAGCAAGCTCTCGGCGCATCTCAAGAGCGCATCAAGAGAGCGCGCATCAACAGCTCTCT 2183
 DB 2076 CTCGAACTCTCAAGAGTCAAGAGCGCATCAAGTCTCGGCGAGAGATCAACAGCTCTCT 2135
 QY 2184 CAAGATGCTCGCGTA 2198
 DB 2136 CAAGATGCTCGCTTA 2150
 RESULT 12
 E01783
 LOCUS E01783 2331 bp DNA linear PAT 29-SEP-1997
 DEFINITION DNA sequence coding for L-phenylalanine ammonialyase.
 ACCESSION E01783
 VERSION E01783.1 GI:2170036
 KEYWORDS JP 1988317086-A/1.
 SOURCE Rhodospiridium toruloides
 ORGANISM Rhodospiridium toruloides
 Bacteria; Fungi; Basidiomycota; Urediniomycetes;
 Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
 1 (bases 1 to 2331)
 Fukuura, N., Yoshino, S., Watanabe, M., Nakajima, Y. and Makiguchi, N.
 RECOMBINANT PLASMID FOR MANIFESTING L-PHENYLALANINE-AMMONIA-DYASE
 AND TRANSFORMANT CONTAINING SAID PLASMID
 Patent: JP 1988317086-A 1 26-DEC-1988;

COMMENT
 MITSUI TOATSU CHEM INC
 OS Rhodospiridium toruloides
 PN JP 1988317086-A/1
 PD 26-DEC-1988
 PF 18-JUN-1987 JP 1987152357
 PI FUKUHARA NOBUHIRO, YOSHINO SADA, WATANABE MIDORI, PI
 NAKAJIMA YOSHIYUKI,
 PI MAKIGUCHI NOBUYOSHI
 PC C12N15/00, C12N1/20//C12N9/88, (C12N1/20, C12R1:01), (C12N9/88, PC
 C12R1:01);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=IPO 559 (ATCC 10788);
 FH Key Location/Qualifiers
 FT CDS 1..2151
 FT 3'UTR /product='L-phenylalanine ammonialyase' FT
 FEATURES
 source Location/Qualifiers
 1..2331 /organism='Rhodospiridium toruloides'
 /mol_type='genomic DNA'
 /db_xref='taxon:5286'
 ORIGIN
 Query Match 49.5%; Score 1196.6; DB 6; Length 2331;
 Best Local Similarity 74.0%; Pred. No. 2,4e-136;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
 99 GCGCGCTCCAGCAAGCGGCTGCGGAGCCCACTTGGGCTTCGCGCGCGAGCCCGGCT 158
 DB 45 GGTGCAATCCGCAAGAGAGCTGCAATGCGGCTTCAACAACTTGCACTGCAAGCTTC 104
 QY 159 CGATGCGCAGCGCGCGCAGCAGTGCAGCTGAGATGTGAGAGAGCTCTGAGGAGCC 218
 DB 105 GCACCTGCCCAACCAAGGTACGAGGTGACATGTGAGAGATGCTGCGCGCGC 164
 QY 219 CAGCGAGAGTGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTG 278
 DB 165 GACCGATTCAGCGCTGAACTGACGCTGCTGCTCAACTGCGAGAGCTGTCTCGCG 224
 QY 279 GCGCGCAGAGGCGGAGGCTCCGCTCCAGAAAGAGAGATCCGCGCAGCGGTGCA 338
 DB 225 CGCGAGAGAGGAGGCGGCTGCTCGGCTCAAGAGAGAGAGAGATCCGCTCAAAATTTGA 284
 QY 339 CAAGAGGTGACTTCTCTCAAGGCTCAGACTGAGTCTGAGTCTGAGAGTACACAGCG 398
 DB 285 CAATGTGTGAGATTTTGTGCTCGCACTTCAAGAGCTGCTGAGAGTCTGAGAGTCTG 344
 QY 399 TTTGAGTGTGCTCGCGCAGACAGAGACTGAGATGAGTCAAGCTTCAAGAGGCGCTCAT 458
 DB 345 ATTGGGAGATCCGAGAGACCGCGACAGAGAGAGAGAGTCTGCTCGAGAGGCTTCTCT 404
 QY 459 CGAGCAGCAAGCTTGGGCGGTGACGCGAGCTGCTGCTTCAAGGCTGAGAGCGCG 518
 DB 405 CGAGCAGCAAGCTTGGGCGGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 464
 QY 519 CTTGAGAGACAGCTTGGCTGAGGTGCTGCGGCGCGCAGTGTCAATCCGCGTCAATC 578
 DB 465 TCTGAGAACTGCTTCCCTGAGGTTTTCGCGCGCGCAGTGTCAATCCGCGTCAATC 524
 QY 579 GCTCAGCGTGGCACTGCGCGCTGCGCTGCTGCTTGAAGGCTCAACAACTTCTT 638
 DB 525 CTTGACCGCGGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
 QY 639 GAACCAAGCATAGCCCATAGTCCCTCGCGGCTTCAATCCGCGTGGCGAGAGCT 698
 DB 585 CAACCAAGCATAGCCCATAGTCCCTCGCGGCTTCAATCCGCGTGGCGAGAGCT 644
 QY 699 GAGCCGCTCTGATCATGCGCGCGCATGACCGGTCAACCGTCAAGTTTCAGT 758

Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCCGACCAATGCGCTGCGGCGCCCACTTCGCGCTCCGCGCACCGCGCCT 158
 Db 45 CGTGCATCCGAAAGCAGCTGTGTAATGGCGCTTCGACCAACCTCGAAGTCGAGCTC 104
 QY 159 CGATGGCCACGCGCGCCACCAAGTCCGAGTCCGAGATCGTGAGAGAGCTCTCAAGCAACC 218
 Db 105 GCACCTGCCCAACCAACCAAGTCCGAGTCCGAGATCGTGAGAGAGAGTCTCGCGCGCC 164
 QY 219 CACCGACGACGTCTGAGAGCTCAGCGGGTACAGCTCCGCTCGGTGAGCTTCGCGCGC 278
 Db 165 GACCGACTGAGCGCTGGAAGTCCGAGTCTCGCTCAACTCCGAGAGCTCTGCTCGGC 224
 QY 279 CGCGCGCAAGGGCGCGAGGCTCGCGTCCAGACGACGAGATCCGCGCACCGCTCGA 338
 Db 225 CGCGAGGAAGGGCGAGGCTCGCGTCCAGAGACGAGATCCGCTCAAGATTGA 284
 QY 339 CAGAGCGCTGCACTTCCTCAAGGCCCACTTCAGAACTCGGTCTTACGAGATCCAGCGG 398
 Db 285 CAATTCGTGAGTCTTCGCTTCGAACTCTCAATGAGCTTACGCGCTCAAGACTGG 344
 QY 399 TTTCGGTCTCGCGCGACACAGAGATGAGAGATGAGTCAAGCTCCAGAGCGCTCAT 458
 Db 345 ATTTGGCGGATCCGAGACACCGGACCGAGAGCGCATTCGCTCCAGAGAGCTCTCT 404
 QY 459 CGAGCACCAAGCTCTCGCGCTGACCGCGACGTCGCTCTCTTCAAGCTCGAGCGCG 518
 Db 405 CGAGCACCAAGCTCTCGCGCTGACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
 QY 519 CCTCGAGAACAGCTTCGCTCGAGTGTCTCGCGCGCGCATGGTCAATCCGCGTCAATC 578
 Db 465 TCTCGAGAACTGCTTCCCTCGAGTGTCTCGCGCGCGCATGGTCAATCCGCGTCAATC 524
 QY 579 GCTCACGCGTGGCCACTCGCGCTGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 638
 Db 525 CTTCGACCGCGCGCGCTCGCGCTGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
 QY 639 GAACACCGCATCAAGCCCATGATCCCTCGCGCGCGCATTCCTCGCGCGCTCGCGCATC 698
 Db 585 CAACCAAGCATCAAGCCCATGATCCCTCGCGCGCGCATTCCTCGCGCGCTCGCGCATC 644
 QY 699 CAGCCGCTCTGATCATGCGCGCGCGCATCAAGCTCAAGCTCAAGCTCAAGCTCAAG 758
 Db 645 CT 704
 QY 759 TTTCGACGAGGGAACCGAGATGATGTTTGGCGCGAGGCGCATCTCTCTCTCTCTCT 818
 Db 705 CATTCAAGAGGGAAGAGATCTGTACGCGCGAGGCGCATCTCTCTCTCTCTCTCTCT 764
 QY 819 CGAGGAGTGTCTCTCGCGCGCGAGAGGATCTGATCTGATCAAGGAGCGCGCATCT 878
 Db 765 CGAGCGCT 824
 QY 879 CGCTCTGATGAGGAGCTCTCAATGATGATGATGATGATGATGATGATGATGATGAT 938
 Db 825 AGCATTCGATGAGGAGCT 884
 QY 939 CTTCGAGGCT 998
 Db 885 GCTTCAGGCGCATGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 944
 QY 999 CCACGAGCTTCGCT 1058
 Db 945 TCACAGCGTCAAGCGCGCT 1004
 QY 1059 CCTTCGCGCT 1118
 Db 1005 CTTGAGGGAAGCGCGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1064
 QY 1119 GGGCATTTCTGCGAGGACGCTACCGGCTCCGACGTCGCTCAAGTCTCTCGCGCGCT 1178
 Db 1065 GGGCATTTCTGCGAGGACGCTACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1124

QY 1179 CGTGGAGACATGATGACGCTTACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
 Db 1125 CGTACGACACTCTTATTCACCGCCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1184
 QY 1236 CGACACCCGCT 1295
 Db 1185 CGACACCCGCT 1244
 QY 1296 GTCCGCT 1355
 Db 1245 TCCGCT 1304
 QY 1356 CAATTCACGAGTGCACCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1415
 Db 1305 CAATTCACGAGTGCACCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1364
 QY 1416 CCTGCGTCCGAGGACCGCT 1475
 Db 1365 CCTGCGTCCGAGGACCGCT 1424
 QY 1476 TGTTCAGCTTCGAGCTCGAGCACTTCGCAACCGGATCACTTCTCTCTCTCTCTCTCT 1535
 Db 1425 GGGTACACCTCGAGTTCGAGCACTTCGCAACCGGATCACTTCTCTCTCTCTCTCTCTCT 1484
 QY 1536 AGAATGGGCAACGAGGCGTCAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
 Db 1485 TGAATGGGCAACGAGGCGTCAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1544
 QY 1596 GGCCACGACGTCCTTCT 1655
 Db 1545 GTCCACGAGTCT 1604
 QY 1656 CGACTTCGCGCGATGAGCTGACCTTCAAGAGAGTTCGACCGCTCTCTCTCTCTCTCTCT 1715
 Db 1605 CGACTTCGCGCGATGAGCTGACCTTCAAGAGAGTTCGACCGCTCTCTCTCTCTCTCTCT 1664
 QY 1716 CTTCCAGACGACCTGCGGCACTGCGCTCGACGCA-----AGCAGCTTGGCGCT 1763
 Db 1665 CATGACCGACGACCTTGGCT 1724
 QY 1764 CGAGTCAAGAGGCGCTCAAGAGGCTTCGAGACGAGACGAGACGAGACGAGACGAGACG 1823
 Db 1725 GAGGTGAAGAGGCGCTCGCAAGGCGCTCGAGAGACGAGACGAGACGAGACGAGACGAG 1784
 QY 1824 GCGCTGCGACGAGGCT 1883
 Db 1785 GCGCTGCGACGAGGCT 1844
 QY 1884 GCGCTTGCACAGTCAACCTTACTGCGGTCGCAACGCGTGAAGTTCCTCGCGCGAGAA 1943
 Db 1845 GTC-----GCTCTGCTGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1895
 QY 1944 GGCATTTGCTCAACGCGCGAGGTCGCAACCGCTTCTGCGAGACGCGCTTCTGCGAGCG 2003
 Db 1896 GGCATTTGCTCAACGCGCGAGGTCGCAACCGCTTCTGCGAGACGCGCTTCTGCGAGCG 1955
 QY 2004 GCGGCGGACGAGTACCTCTCGCGCGGACGCGCGCTGATCTGCTCTCTCTCTCTCTCTCT 2063
 Db 1956 GCGGCGCTTCTGATCTCTCGCGCGGACGCGCGCTGATCTCTCTCTCTCTCTCTCTCTCT 2015
 QY 2064 GCTCGGCTGACAGGCGCGCGCGAGCGTGTCTGCGCTGCGACGAGACGAGACGAGTCTG 2123
 Db 2015 GCTTGGGCTCAAGGCGCGCGCGAGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2075
 QY 2124 GAGCAAGTCTCGCGCATCTAGAGGCTCTAGAGGCGCGCGCATCTCTCTCTCTCTCTCTCT 2183
 Db 2076 CTGCAAGTCTCTCAAGATCTAGAGGCTCTAGAGGCGCGCATCTCTCTCTCTCTCTCT 2135
 QY 2184 CAAGATGCTGCGTA 2198
 Db 2136 CAAGATGCTGCTTA 2150

RESULT 14

LOCUS E01785 2331 bp DNA linear PAT 29-SEP-1997

DEFINITION DNA sequence coding for L-phenylalanine ammonia-lyase.

ACCESSION E01785

VERSION E01785.1 GI:2170038

KEYWORDS JP 1988317088-A/1.

SOURCE Rhodospiridium toruloides

ORGANISM Rhodospiridium toruloides

Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Microbotryomycetidae; Sporidiobolales; Rhodospiridium.

1 (bases 1 to 2331)

Fukuhara,N., Yoshino,S., Watanabe,M., Yamamoto,K., Suzuki,M. and Nakajima,Y.

CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF ADVENTITIOUS GENE PRODUCT USING SAID CONTROL

Patent: JP 1988317088-A 1 26-DEC-1988;

JOURNAL MITSUI TOATSU CHEM INC

OS Rhodospiridium toruloides

PN JP 1988317088-A/1

PD 26-DEC-1988

PF 18-JUN-1987 JP 1987152359

PI FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI YAMAMOTO KAORU,

PI SUZUKI MAKI, NAKAJIMA YOSHIYUKI

PC C12N15/00,C12N9/88,(C12N9/88,C12R1:01);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC key *source: strain=IFO 559 (ATCC 10788);

PH Location/Qualifiers

FT CDS 1..2151

FT 3'UTR 2152..2331.

FEATURES

source Location/Qualifiers

1..2331

/organism="Rhodospiridium toruloides"

/mol_type="genomic DNA"

/db_xref="taxon:5286"

ORIGIN

Query Match 49.5%; Score 1196.6; DB 6; Length 2331;

Best Local Similarity 74.0%; Pred. No. 2.4e-136;

Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

99 GCGCGCTCCGACCAAGTGGCGTGGCGGCCGCACTTCGGCTCTCGCGGCAAGCCGGCCT 158

45 CTTCCGATCCGTAAGCAAGGCTGTCAATGGCGCCTCGACCACTTCGAGTCCGAGGCTC 104

159 CGATGGCCAGCGCCGGGCAACAGTCGACAGCTTGAGATCGTGCAGAGCTCTTCAAGCAACC 218

105 GCACTGCCCAACCAACCGAGTCAAGAGTCAATGTCAGAAAGATGCTCGCGCGGCC 164

219 CACCGACAGCTGTCTGAGCTTACGGGCTACAGCTTACCGTCCGTACAGTGTTCGGCGC 278

165 GACCACTCGACGCTGCACTGACGCGTACTCGCTCACTCGAAGAGTGTGCTTCGCGC 224

279 CCGCCGCAAGGGGCGCAGAGGTCGCGCTCCAGAAAGACGACGAGATCGCGCAAGCTTCGA 338

225 CCGGAGAAAGGCGAGGCTGTCCGCTCAAGAGACGACGAGATCCGCTCAAGATGTA 284

339 CAAGAGCGTGACTTCTTCAAGGCCCAAGCTTCAAGACTCGTCTTACGAGAGTCAACACGCG 358

285 CAAATCGGTGAGGTTCTTCCGCTCGCACTCTCCATGAGAGTCTTACGCGCTCAAGATGG 344

399 TTTCGGTGTCTGGCGGACAGAGAGTCAAGAGATGAGTCAAGCTTCAAGAGGCGCTCAT 458

345 ATTGGCGATCGAGACACCGGACAGAGAGACCAATTCGCTTCCAGAAAGGCTTCTCT 404

459 CGAGCACCAAGCTCTGGGCGTGACCGCAGAGTCCGCTCGTCTTCAAGCTTCGAGCGCG 518

405 CGAGCACCAAGCTCTGGGCGTGTCTCCCTTGTGCTTCACTGCTTCGGCTCGGCGCGCG 464

519 CTTGAGAAACAGCTTCCCTTCGAGGTCGTCGCGGCGGCGGCAATGATCATCGGTCACACTC 578

465 TCTGAGAACTCGGTTCTCCCTTCAGAGTGTTCGCGGCGGCGGCAATGATCATCGGTCACACTC 524

579 GCTCAGCGTGTGACACTCGGCGGTCGCTCTGCTCTGCTTGAAGCGCTCAACCACTTCTT 638

525 CTTGACCCGGGCGCACTCGGCTGTCCGCTCTGCTCTGAGGCGCTCAACCACTTCTT 584

639 GAACACCGGATACCGGCACTGCTCCCTTCGCGGCTCACTTCGCGGTCGCGGCACT 698

585 CAACACCGGATACCGGCACTGCTCCCTTCGCGGCGGCGGCACTTCGCGTTCGCGGCACT 644

699 CAGCCGCTCTGATCATCTCCGCGGCACTACCGGTCACCGGCAAGTCAAGTTCAGCT 758

645 CTTCTCTCTCTCTCACTTGAAGCGGCACTACCGGTCACCGGCAAGTTCAGCT 704

759 TTGACAGAGGAAACGAGAAATCATGTTTGCAGCGGAGGCACTTCGCTTTTGTCT 818

705 CGTCCAGAGGCAAGAGAAATCTGTACGCGCGGCGGCAAGTTCGCTTTCAACT 764

819 CGAGGCACTGCTCTCGGCGGCGGCAAGAGGCTCGGCTGTGTCAGAGGCAAGGCGCTCTC 878

765 CGAGCGGTCGCTCTCGGCGGCGGCAAGAGGCTCTGGTCTGTGTCAGAGGCAAGGCGCTCTC 824

879 CGCCTGATGCGACCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 938

825 AGCATGATGCGACCTCTCGGTCGAGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 884

939 CTTGACGCTCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 998

885 GCTCAGCGGCAAGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 944

999 CCAGCAGCTGTCCGCGGCGGCAAGGCGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1058

945 TCAGACGTCACGCGGCTCAAGGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1004

1059 CTTTCGCGCTGTGCTTGTGCGGTCGAGCAAGAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1118

1005 CTTGAGAGGAAACGCGCTTGTGCTGTCACATGAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1064

1119 GGGATTTTGGCCAGAGACGCTACCGGCTCCGAGAGTGGCTCACTTCTCGGCGGCT 1178

1065 GGGATTTTGGCCAGAGACGCTACCGGCTCCGAGAGTGGCTCACTTCTCGGCGGCT 1124

1179 CGTGAAGACATGATGACGCTTACTGACTCTCTGCTGAGAAACA--CGACGAC 1235

1125 CGTGAAGACATGATGACGCTTACTGACTCTCTGACTGAGAAACA--CGACGAC 1184

1236 CGACACCGGCTCTCTGACGCTGAGAAACAAGCAGACCGGCGGCAACTTCCAGAGC 1295

1185 CGACACCGGCTCTCTGACGCTGAGAAACAAGCAGACCGGCGGCAACTTCCAGAGC 1244

1296 GTCCGCTGTCTGATTTGATGAGAAAGACAGGCTGCACTGCGCTTCAAGAGCT 1355

1245 TGCCGCTGTGACCAACACATGAGAAAGACCTCGCTGCGCTGCGGCAAGCT 1304

1356 CAATTCAAGAGTCAAGGCTTCTGACGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1415

1305 CAATTCAAGAGTCAAGGCTTCTGACGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1364

1416 CTTGCGGCGGAGAGACCGGCTGCTCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1475

1365 CTTGCGGCGGAGAGACCGGCTGCTGCTTCTTCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 1424

1476 TGCTTACCTTTCAGAGTTCGCGCACTTTCGCAACCGGCTCACTTCTGTCAGCGCGC 1535

1425 GAGCGTACACTTCGAGGTTGGGACACTTCGCAACCGGCTGCACTTTCAGCGCGC 1484

1536 AGAGTGGCAACAGGCGGCTCACTGCTGCTTCTTCACTTTCGCGGCGGCACTGCGCA 1595

Db 1125 GGTCAAGCACTTCAACGCCCAAGCCCTCTCACCATCGAGCCGGCCAGTCAGACAC 1184
2Y 1236 CGACAAACCCGCTCTCTCGACGTGAGAAACAAGACAACGCGACGCGGCAACTTCAGGC 1295
Db 1185 CGACAAACCCCTCTCAACGACGTGAGAAACAAGACTTCGACCAACGCGGCAATTCCAGGC 1244
2Y 1236 GTGGCTGTCTCGATTTCGATGAGAAAGACCAAGGCTCGCACTCGCCCTCATTCGGCAAGCT 1355
Db 1245 TCCGCTGTGGCAACACATGAGAAAGACTCGCTTGGGCTCGCCCAATTCGCAAGCT 1304
2Y 1356 CAACCTTCACGAGTGCACCGAGTGTCTCAACGTGCCATGAAACCGGAGCTGCTTCGTG 1415
Db 1305 CAACCTTCACGAGTGCACCGAGTGTCTCAACGCGGATGAAACCGGAGCTTCCTTCGTG 1364
2Y 1416 CCTGCTGCCGAGGACCCGCTCGTCAACTATACGCGCAAGGCTTGGACATTCACATCGC 1475
Db 1365 CCTCGCGGCCGAAAGCCCTCGCTCTCTCACTGCAAGGAGGCTCGACATTCGCGCTGC 1424
2Y 1476 TGCTTAAGCTTCGAGAGCTCGGCGCACTTGGCAACCGGATCACTACTTCGTCAGCCCGC 1535
Db 1425 GAGCTACACCTCGAGATTGGAGACACCTCGCAACCTGTGACGAGCATTCAGCCGCGC 1484
2Y 1536 AGAGATGGGCAACCAAGCCGTCAACTGCTGCTCATCTCGCGGCGCGCACTGCCGA 1595
Db 1485 TGAGATGGCGAAACCAAGGCGTCAACTGCTTGCCTCATCTCGGCTGTGCAACCGA 1544
2Y 1536 GGCACAGACGCTCTTTCT 1655
Db 1545 GTCCACACGAGCTCTTCT 1604
2Y 1656 CGACCTCGCGCGATGAGAGCTCGACTTCAGAAAGAGATTCGACCCGCTTCTCCGACTCT 1715
Db 1605 CGACTTGGCGCGATCGAGTTCGAGTTCAAGAAAGATTCGCGCCAGCCATTCGTCTGCT 1664
2Y 1716 CCTCAGAGAGACCTCGGCACTGAGCTGACGTCA-----ACGCACTTGGCT 1763
Db 1665 CATGACCAAGACATTGGCTCGGCATGACCGGCTCGAACCTGCGCGACGAGCTGTGCA 1724
2Y 1764 CGAGGTCAAGAAAGGCGTCAACAAGCGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
Db 1725 GAAGGTGAACAAAGAGCTTCGCGAAGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAG 1784
2Y 1824 GCGCTGAGACGACGCTTCTGTGTAAGCAGCCGAGACCGTTCGTGAGCTCCTCTGTCCTC 1883
Db 1785 GCGCTGGACGACGCTTCTCTCTGCGCGGACCGTTCGTGAGAGTCTCTGTCGAC 1844
2Y 1884 GCGCTTGGCAACGTACCTTACTGCGCGTCAACGCGTGAAGGTTCCTCGGCGAGAA 1943
Db 1845 GTC-----GCTCTGCTTCGCGCGCTCAACGCTGGAAGTTCGCGCGGAGTC 1895
2Y 1944 GGCATCTCGCTCAAGCGAGGAGTGCAGACCGCTTCGAGAGAGCGCGCTTCGAGGC 2003
Db 1896 GGCATCTCGCTCAACCGCGCAAGTCCGAGAGACTTCGTGCGCGCGCTGACCTGCTC 1955
2Y 2004 GCGGCGGACGCGATACCTTCGCGCGGAGACGCGCTTCGTGCTGTCGTGCGGAGGA 2063
Db 1956 GCGCGGCTCTGTGTACTTCGCGCGGACCTCAAGATCCTTCAACGCTTCGTCGCGAGGA 2015
2Y 2064 GCTCGCGGTGAGAGGCGCGCGCGGAGAGTGTTCGTGCGCTGAGAGAGAGAGAGATCGG 2123
Db 2016 GCTTGGCGTCAAGGCGCGCGCGGAGAGTCTTCTCGCAAGAGAGAGAGATCGG 2075
2Y 2124 GAGCAAGCTTCGCGCATTCAGAGGCGATCAAGAGACGCGCGCATCAACAGCTCTGCT 2183
Db 2076 CTCGAAAGCTTCGAAGATTCAGAGGCGATCAAGTCGCGGAGATCAACAGCTCTCTCT 2135
2Y 2184 CAAGATGCTCGCGTA 2198
Db 2136 CAAGATGCTCGCTTA 2150

Dp	721	GGCCCAATCAACCGGATCCGCCGACGTCGAAGTTTGCACGAGGGAACCGAANAAG	780
Qy	781	ATCATGTTTGGCGGCGGACAGGCAATCTGGCTCTTTGGTCTTCGAGGACGATGCTCTTGCCCG	840
Dp	781	ATCATGTTTGGCGGCGGACAGGCAATCTGGCTCTTTGGTCTTCGAGGACGATGCTCTTGCCCG	840
Qy	841	AAGAGGGATCTCGGTCTGGTCAACGGAACGACCGCTCTCCGCTCGATAGAGCAACCTCAGT	900
Dp	841	AAGAGGGATCTCGGTCTGGTCAACGGAACGACCGCTCTCCGCTCGATAGAGCAACCTCAGT	900
Qy	901	CTGACACGATCTCGACATGCTCTTGCTCTCTTCGCAAGGCTTTGACGGCTTTCACGGTGGAG	960
Dp	901	CTGACACGATCTCGACATGCTCTTGCTCTCTTCGCAAGGCTTTGACGGCTTTCACGGTGGAG	960
Qy	961	GCCATGTGTCGAGCAGAGGAGCTGTTCGAGCGCTTTCATTCACGACGCTCCGCCCGGAC	1020
Dp	961	GCCATGTGTCGAGCAGAGGAGCTGTTCGAGCGCTTTCATTCACGACGCTCCGCCCGGAC	1020
Qy	1021	CCCGGCCAGGTCGAGGTCGCGGCAACATCCGCAAGCTCTTTCGGCTGCTGTTTGGC	1080
Dp	1021	CCCGGCCAGGTCGAGGTCGCGGCAACATCCGCAAGCTCTTTCGGCTGCTGTTTGGC	1080
Qy	1081	GTTGAGACGAGGAGGAGGATCAAGTTCAGACACGACGAGGCAATTTCTTCCGACGACCG	1140
Dp	1081	GTTGAGACGAGGAGGAGGATCAAGTTCAGACACGACGAGGCAATTTCTTCCGACGACCG	1140
Qy	1141	TACCCGCTCCGCAAGTCGAGTCTCAGTTCTCGGCGGCTGTGAGAGCATATGACGCGC	1200
Dp	1141	TACCCGCTCCGCAAGTCGAGTCTCAGTTCTCGGCGGCTGTGAGAGCATATGACGCGC	1200
Qy	1201	TACTCGACCTCTCTCGCTCGAAGAACACGACGACCGAACACCCGCTCTTCGACGTGAG	1260
Dp	1201	TACTCGACCTCTCTCGCTCGAAGAACACGACGACCGAACACCCGCTCTTCGACGTGAG	1260
Qy	1261	AACAGAGAGACCGCGACCGCGGCAACTTCCAGGAGTCGAGTGTCTCGATTTGATGAGAG	1320
Dp	1261	AACAGAGAGACCGCGACCGCGGCAACTTCCAGGAGTCGAGTGTCTCGATTTGATGAGAG	1320
Qy	1321	AAGACCAAGGTCGCACTCGCCTCATCGGCAAGCTCAATTCAACGACAGTCACGAGTTG	1380
Dp	1321	AAGACCAAGGTCGCACTCGCCTCATCGGCAAGCTCAATTCAACGACAGTCACGAGTTG	1380
Qy	1381	CTGAACGCTGCATGAACCGGGGCGCTGCTGTGGCTCGCTGCGAGAACCCGTCGCTC	1440
Dp	1381	CTGAACGCTGCATGAACCGGGGCGCTGCTGTGGCTCGCTGCGAGAACCCGTCGCTC	1440
Qy	1441	AACATACACGAGGAGGCTTGACATTGACATCGCTCTTACGCTTCGAGGCTCGAGCAC	1500
Dp	1441	AACATACACGAGGAGGCTTGAGACATTGACATCGCTCTTACGCTTCGAGGCTCGAGCAC	1500
Qy	1501	CTTGGCAACCGCGTCACTACCTTGTGCCAGCCGCGACAGATGGGCAACAGGCGCTCAAC	1560
Dp	1501	CTTGGCAACCGCGTCACTACCTTGTGCCAGCCGCGACAGATGGGCAACAGGCGCTCAAC	1560
Qy	1561	TGCGCTGCTCATCTTCGCGGCGCGACCTGACGAGGCAACAGACGCTCTTCTCGCTT	1620
Dp	1561	TGCGCTGCTCATCTTCGCGGCGCGACCTGACGAGGCAACAGACGCTCTTCTCGCTT	1620
Qy	1621	CTGCGCTCGACACTGTATCTGACGCTTCAGGCGGTGACCTTCGCGCGATGAGAGCTGCAC	1680
Dp	1621	CTGCGCTCGACACTGTATCTGACGCTTCAGGCGGTGACCTTCGCGCGATGAGAGCTGCAC	1680
Qy	1681	TTCAAGAAACAGTTTGACACCGCGCTTCTCCGACCTCTCTCAAGAGACCTCGGACCTGAC	1740
Dp	1681	TTCAAGAAACAGTTTGACACCGCGCTTCTCCGACCTCTCTCAAGAGACCTCGGACCTGAC	1740
Qy	1741	CTCGACGTCACGCACTTGCCTCGAGGTCAGAGGCGCTCAACAGGCTCTCGACGAG	1800
Dp	1741	CTCGACGTCACGCACTTGCCTCGAGGTCAGAGGCGCTCAACAGGCTCTCGACGAG	1800
Qy	1801	ACGACGACGTCACGCTTCGAGCGCGGCTGGACGAGACGCTTCTCGATCGGACCGGACCC	1860
Dp	1801	ACGACGACGTCACGCTTCGAGCGCGGCTGGACGAGACGCTTCTCGATCGGACCGGACCC	1860

QY	1861	GTGGTGAAGACTCTCTCTGCTCTCGCCCTCGCCCTGACAGTACACCTTATCTGCGCTCAAGCG	1920
Db	1861	GTGGTGAAGACTCTCTCTGCTCTCGCCCTCGCCCTGACAGTACACCTTATCTGCGCTCAAGCG <td>1920</td>	1920
QY	1921	TGGAAGTGTGCTCGGGCGGAAGAGCATCTCGCTCAAGCGAGTGTGCGCAACCGCTTC	1980
Db	1921	TGGAAGTGTGCTCGGGCGGAAGAGCATCTCGCTCAAGCGAGTGTGCGCAACCGCTTC	1980
QY	1981	TGSCAGACGCGCTTCTGCGCAGCGCGCGAGCACTACCTCTGCGCGCGAGCAAGCGCGCTC	2040
Db	1981	TGSCAGACGCGCTTCTGCGCAGCGCGCGAGCACTACCTCTGCGCGCGAGCAAGCGCGCTC	2040
QY	2041	CTGTACTGCTTGTGTCGCGAGAGACTCGGCGTCAAGCGCGCGCGCGAGCGCTTTTGTG	2100
Db	2041	CTGTACTGCTTGTGTCGCGAGAGACTCGGCGTCAAGCGCGCGCGCGAGCGCTTTTGTG	2100
QY	2101	GGGCGGACGAGAGAGATCGGAGAGCACTCTGCGCGATCTACAGAGCCATCAAGAGC	2160
Db	2101	GGGCGGACGAGAGAGATCGGAGAGCACTCTGCGCGATCTACAGAGCCATCAAGAGC	2160
QY	2161	GGCGCGATCAACCAAGTCTCTGTCAGATCTCGCGCTAAGAGCGCGAGAGCGCTGCGCTA	2220
Db	2161	GGCGCGATCAACCAAGTCTCTGTCAGATCTCGCGCTAAGAGCGCGAGAGCGCTGCGCTA	2220
QY	2221	GAGCGCGCGCTCACCCCAAGACCAAGCTTTTCCACGTCGTGTGTGCCAAGAACGACCTT	2280
Db	2221	GAGCGCGCGCTCACCCCAAGACCAAGCTTTTCCACGTCGTGTGTGCCAAGAACGACCTT	2280
QY	2281	CCTTCATACAGATGCGCTTACTCTCTGCGCGCTACAGCTCTCAGTCTTTTGGTAT	2340
Db	2281	CCTTCATACAGATGCGCTTACTCTCTGCGCGCTACAGCTCTCAGTCTTTTGGTAT	2340
QY	2341	CCCCGCTCTCTCGGTCGTCAGTACAGTGTATAGACCTGGAATGGATTGCAAGTCTTCG	2400
Db	2341	CCCCGCTCTCTCGGTCGTCAGTACAGTGTATAGACCTGGAATGGATTGCAAGTCTTCG	2400
QY	2401	AGTTCAAAAAAGAAAAA 2419	
Db	2401	AGTTCAAAAAAGAAAAA 2419	
RESULT 3			
ID	ABA95251	standard; DNA; 2741 BP.	
AC	ABA95251;		
XX	XX		
DT	10-JUN-2002	(first entry)	
XX	XX		
DE	R. graminis PAL genomic DNA sequence.		
XX	XX		
KM	PAL; yeast; phenylalanine ammonia lyase; cinamic acid; cinamate; ds;		
XX	cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.		
OS	Rhodocortula graminis.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	1..2589	
FT	FT	/*cag= d	
FT	FT	/product= "PAL polypeptide"	
FT	FT	/note= "contains introns"	
FT	FT	1..361	
FT	FT	/*cag= c	
FT	FT	/number= 1	
FT	FT	6	
FT	FT	misc_feature	
FT	FT	/*cag= b	
FT	FT	/note= "PAL DNA of strain ATCC 20804 at this position	
FT	FT	contains the nucleotide C and the PAL DNA of strain KGX	
FT	FT	362..448	
FT	FT	/*cag= d	
FT	FT	/number= 1	

443. .880
/*tag= e
/number= 2
881. .960
/*tag= f
/number= 2
961. .1295
/*tag= g
/number= 3
1296. .1364
/*tag= h
/number= 3
1365. .1529
/*tag= i
/number= 4
1530. .1586
/*tag= j
/number= 4
1587. .1748
/*tag= k
/number= 5
1749. .1821
/*tag= l
/number= 5
1822. .1947
/*tag= m
/number= 6
1948. .2007
/*tag= n
/number= 6
2008. .2586
/*tag= o
/number= 7
WO200208402-A2.
31-JAN-2002.
24-JUL-2001; 2001WO-US023270.
24-JUL-2000; 2000US-00624693.
(PCBU-) PCBU SERVICES INC.
Yoshida RK, Koetsura AB;
WPI: 2002-268973/31.
P-PsDB; ABB07690.
Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
treating mammal having disease or disorder from phenylketonuria, cancer,
human immunodeficiency virus infection and human cytomegalovirus
infection.
Claim 4, Fig 3, 4, 135pp; English.
The invention relates to an isolated and purified yeast phenylalanine
ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
host cell comprising the polynucleotide is useful for the production of L
-PAL. By adding the polynucleotide or the host cell to a composition
comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
improving the production of PAL, its analog or another optically active
unnatural amino acid having PAL-like structure. The PAL polynucleotides
are useful for treating a mammal having a disease, disorder or condition
selected from phenylketonuria, cancer, human immunodeficiency virus
infection and human cytomegalovirus infection. The present sequence
represents a R. graminis PAL genomic DNA sequence
Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;
Query Match 75.5%; Score 1826.6; DB 6; Length 2741;
Best Local Similarity 84.3%; Pred. No. 1.5e-280;
Matches 2312; Conservative 2; Mismatches 1; Indels 426; Gaps 6;

QY 37 ATGGCCCCCTTCCTTGAGCTCGCTCGCCACCAAGCTGCTTACCAACGGCTCG 96
DB 1 ATGGCMCTTCTCTTGAGCTCGCTCGCCACCAAGCTGCTTACCAACGGCTCG 60
QY 97 CAGCGCGCTCCGACCAAGTGGCTGGGAGCCCACTTGGGCTCTCCGCGGACGCGCGG 156
DB 61 CAGCGCGCTCCGACCAAGTGGCTGGGAGCCCACTTGGGCTCTCCGCGGACGCGCGG 120
QY 157 CTGGATGGCCAGCGCGGCGGACCAAGTGGAGCTCGAGATGGAGAGGCTCTCAGCGAC 216
DB 121 CTGGATGGCCAGCGCGGCGGACCAAGTGGAGCTCGAGATGGAGAGGCTCTCAGCGAC 180
QY 217 CCAACGACGACGCTGCTGAGCTCAGCGGGTACAGCTCAGCTCGGTGCTGGCT 276
DB 181 CCAACGACGACGCTGCTGAGCTCAGCGGGTACAGCTCAGCTCGGTGCTGGCT 240
QY 277 GCGCGCGGCAAGGGGCGGAGGCTCGCGCTCCAGAAAGAGAGAGATCGCGGCGGCTG 336
DB 241 GCGCGCGGCAAGGGGCGGAGGCTCGCGCTCCAGAAAGAGAGATCGCGGCGGCTG 300
QY 337 GACAAGAGCGTGCACCTTCTCAAGGCGGAGCTTCAAGACTCGGTCTTACGAGTCAACAC 395
DB 301 GACAAGAGCGTGCACCTTCTCAAGGCGGAGCTTCAAGACTCGGTCTTACGAGTCAACAC 360
QY 396 ----- 395
DB 361 GTGCGTTCCGAGACGAGAGGGGAGAACTCGGAGTGCAGGCGCTGAAAGCTGACATTC 420
QY 396 -----GGTTTCGTGGCTCGGCGGACGAGAGCTGAG 429
DB 421 GCTTGAAGGCTGCGCGGCTGCTTGCAAGGTTTCGTGGCTCGGCGGACGAGAGCTGAG 480
QY 430 GATGAGTCAAGCTTCCAGAAAGGCGCTCATGAGACACAGCTTGGCGGCTGACGCGGAG 489
DB 481 GATGAGTCAAGCTTCCAGAAAGGCGCTCATGAGACACAGCTTGGCGGCTGACGCGGAG 540
QY 490 TCGGCTCGCTTCCAGCGTCCGAGCGGCGCTCGAGAAACAGCTTCCGCTCGAGTCTGTC 549
DB 541 TCGGCTCGCTTCCAGCGTCCGAGCGGCGCTCGAGAAACAGCTTCCGCTCGAGTCTGTC 600
QY 550 CGCGCGCATGATGATCCGAGGCTCATGAGCTGAGCGGAGGAGCTTGGCGGCTGTC 609
DB 601 CGCGCGCATGATGATCCGAGGCTCATGAGCTGAGCGGAGGAGCTTGGCGGCTGTC 660
QY 610 GTGCTCTTGAAGGCGCTGACCACTTCTTGAACCAAGGATCAAGCGGATCGTCCGCTC 669
DB 661 GTGCTCTTGAAGGCGCTGACCACTTCTTGAACCAAGGATCAAGCGGATCGTCCGCTC 720
QY 670 CGCGGCTCATGATGATCCGAGGCTCATGAGCTGAGCGGAGGAGCTTGGCGGCTGTC 729
DB 721 CGCGGCTCATGATGATCCGAGGCTCATGAGCTGAGCGGAGGAGCTTGGCGGCTGTC 780
QY 730 ACCGGTCAACCCGAGAGTCAAGGTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 789
DB 781 ACCGGTCAACCCGAGAGTCAAGGTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
QY 790 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 828
DB 841 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
QY 829 -----G 829
DB 901 GAGAGAGCTGTGAGAGGAGTCCCAAGTTTCTGAGTCCGCTTGTTCATGCGATTGAG 960
QY 830 TCTTGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 889
DB 961 TCTTGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
QY 890 CGAGCCTGAGTGTGACAGCTGACAGATCTGCTCTCTGACAGGCTTGAAGGCTC 949
DB 1021 CGAGCCTGAGTGTGACAGCTGACAGATCTGCTCTCTGACAGGCTTGAAGGCTC 1080

FT exon 1355..1529
 FT /tag= h
 FT intron 1530..1586
 FT /tag= 1
 FT exon 1587..1748
 FT /tag= j
 FT intron 1749..1821
 FT /tag= k
 FT exon 1822..1947
 FT /tag= l
 FT intron 1948..2007
 FT /tag= m
 FT exon 2008..2589
 FT /tag= n
 FT terminator 2587..2589
 FT /tag= o

US2002102712-A1.

01-AUG-2002.

24-AUG-2001; 2001US-00939408.

24-JUL-2000; 2000US-00624693.

24-JUL-2001; 2001WO-US023270.

(PCBU-) PCBU SERVICES INC.

Yoshida RK, Kootstra AB;

WPI; 2002-690616/74.

Novel isolated and purified Rhodotorula phenylalanine ammonia lyase polypeptide, useful for treating a mammal having phenylketonuria, cancer,

human immunodeficiency virus or human cytomagalovirus infection.

Claim 4; Page 50-53; 74p; English.

The present invention relates to yeast (e.g. Rhodotorula) phenylalanine ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding such proteins. PAL sequences are useful for producing L-phenylalanine. They are useful for treating mammals having diseases, disorders or conditions that would benefit from treatment with PAL proteins such as phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomagalovirus (HCMV) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. graminis PAL genomic DNA

Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;

Query Match 75.5%; Score 1826.6; DB 6; Length 2741;

Best Local Similarity 84.3%; Pred. No. 1.5e-280; Matches 2312; Conservative 2; Mismatches 1; Indels 426; Gaps 6;

QY 37 ATGGCCCTTCTTGAGCTGCTGCGCACACGCTGCGCAACGCGTTTACCAAGGCTCG 96
 DB 1 ATGGCCCTTCTTGAGCTGCTGCGCACACGCTGCGCAACGCGTTTACCAAGGCTCG 60
 QY 97 CAGCGCGCTCCACCAAGTGGGCTGGGGGCCCACTTGGGCTTCCGCGGACGCCGCGC 156
 DB 61 CAGCGCGCTCCACCAAGTGGGCTGGGGGCCCACTTGGGCTTCCGCGGACGCCGCGC 120
 QY 157 CTGATGAGCAGCGCGGACGACAGTGCAGTGCAGATGTGTCAGAGACTCTCAGCGAC 216
 DB 121 CTGATGAGCAGCGCGGACGACAGTGCAGTGCAGATGTGTCAGAGACTCTCAGCGAC 180
 QY 217 CCCACGACGAGCTGCTGAGCTCAGCGGGTACAGCTTACCGTTCGGTTCGGC 276
 DB 181 CCCACGACGAGCTGCTGAGCTCAGCGGGTACAGCTTACCGTTCGGTTCGGC 240
 QY 277 GCGCGCGGAGGGGGGAGGAGTCCGCTCCAGAAAGAGAGAGAGATCGGCGCAGCGCTC 336
 DB 241 GCGCGCGGAGGGGGGAGGAGTCCGCTCCAGAAAGAGAGAGAGATCGGCGCAGCGCTC 300

QY 337 GACAAGAGCGTGCAGCTTCTCAAGGCCAGCTTCAGAACTCGGTTCACGAGTCAACAC- 395
 DB 301 GACAAGAGCGTGCAGCTTCTCAAGGCCAGCTTCAGAACTCGGTTCACGAGTCAACACAG 360
 QY 396 ----- 395
 DB 361 GTGCGTTCGAGACGAGAGCGGAAATCTGGAGTCCGACGCTGAAACGCTGACATTC 420
 QY 396 -----GGTTTCGTGGCTCGGCCGACACAGAGACTGAG 429
 DB 421 GCTTGAACGGCTCGCGGCTTTCGAGAGGTTTCGTGGCTCGGCCGACACAGAGACTGAG 480
 QY 430 GATCAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGGAG 540
 DB 481 GATCAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGGAG 540
 QY 490 TCCGCTCTGCTTCAAGCTTCGAGAGCGGCGCTCGAAGAACAGCTTCGCTCGAGGTCGTC 549
 DB 541 TCCGCTCTGCTTCAAGCTTCGAGAGCGGCGCTCGAAGAACAGCTTCGCTCGAGGTCGTC 600
 QY 550 CGCGAGCGCATGCTCATCGGCTCACTCGCTCAAGCGGCTCGAGCTCGCGCTC 609
 DB 601 CGCGAGCGCATGCTCATCGGCTCACTCGCTCAAGCGGCTCGAGCTCGCGCTC 660
 QY 610 GTCTCTTGAAGCGCTCAACCACTTTGAACACCGCATCACGCCCATCTCTCCCTC 669
 DB 661 GTCTCTTGAAGCGCTCAACCACTTTGAACACCGCATCACGCCCATCTCTCCCTC 720
 QY 670 CGCGCTTCATCTCGGCTCGGCGGAGCTCAGCGGCTCGATCATCGCGGCGCATC 729
 DB 721 CGCGCTTCATCTCGGCTCGGCGGAGCTCAGCGGCTCGATCATCGCGGCGCATC 780
 QY 730 ACCGCTCAACCGCGAGCTCAAGTTCAGGTTTGAACGAGGAAACGAGATCATGTTT 789
 DB 781 ACCGCTCAACCGCGAGCTCAAGTTCAGGTTTGAACGAGGAAACGAGATCATGTTT 840
 QY 790 GCGCGGAGGCTATCTGCTCTTTGGTCTGAGGCACTC----- 828
 DB 841 GCGCGGAGGCTATCTGCTCTTTGGTCTGAGGCACTGCTGCGAGTCTGCGAGTCTGACT 900
 QY 829 -----G 829
 DB 901 GCAGTGAAGCTTTCAGAGTCTCCCAAGTTTCTGCTACGCTCTTGTTCATGCGATTGAG 960
 QY 830 TCTTGGGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 889
 DB 961 TCTTGGGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 QY 890 CGAGCTCATGCTGAGAGCTGCAATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 949
 DB 1021 CGAGCTCATGCTGAGAGCTGCAATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 QY 950 TCACGTTGAAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1009
 DB 1081 TCACGTTGAAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 QY 1010 GCGCGCGGACCCCGGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1069
 DB 1141 GCGCGCGGACCCCGGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1200
 QY 1070 CGTCTGTTGCGGTTGAGAGACGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAG 1129
 DB 1201 CGTCTGTTGCGGTTGAGAGACGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAG 1260
 QY 1130 GCGAGAGCGCTATCCCGCTCGCAGTGCCT-----CAGTTCTCGCGCGCTCG 1161
 DB 1261 GCGAGAGCGCTATCCCGCTCGCAGTGCCT-----CAGTTCTCGCGCGCTCG 1320
 QY 1162 -----CAGTTCTCGCGCGCTCG 1180
 DB 1321 CTTCGCTCGACCGGCGCTGAGACTTACGTTTTCGATTCAGTTCTCGCGCGCTCG 1380

QY	1181	TGAGAGACATGATGACACGGCTACTGACCTCTTCGCTGAGAACACACGACGACACA	1240
Db	1381	TGAGAGACATGATGACACGGCTACTGACCTCTTCGCTGAGAACACACGACGACACA	1440
QY	1241	ACCGCGCTCCGACGCTCGAGAACCAAGACACACCGGCGACCGGCGCAACTCCAGCGCTCGG	1300
Db	1441	ACCGCGCTCCGACGCTCGAGAACCAAGACACCGGCGACCGGCGCAACTCCAGCGCTCGG	1500
QY	1301	CTGTCTCGATTTTGATGGAAGAC-----	1325
Db	1501	CTGTCTCGATTTTGATGGAAGACAGAGTGGCTCTCGCTGCGCTTCTGTAACGATCT	1560
QY	1326	-----CAGGCTCGACCTGGCCCTATGGGACAGGTCAACTTCA	1363
Db	1561	TGTGCTGAATGTTCTTCTTCCTCGAAGCTGGACTGGCCCTATGGGACAGGTCAACTTCA	1620
QY	1364	CGGAGTGACCGAGTTTGCTCAACGCTGGCCATGAACCGCGGCGCTGTTCTGTCGCTGCTG	1423
Db	1621	CGGAGTGACCGGAGTTTGCTCAACGCTGGCCATGAACCGCGGCGCTGTTCTGTCGCTGCTG	1680
QY	1424	CCGAGGACCCGCTCGCTCAACTATACCGGCAAGGCTTGGACATTCAATCATGCTGCTTAACG	1483
Db	1681	CCGAGGACCCGCTCGCTCAACTATACCGGCAAGGCTTGGACATTCAATCATGCTGCTTAACG	1740
QY	1484	CTTCGG-----	1489
Db	1741	CTTCGGAGGTGAGCGCTGACGTTCTCGCGCGTGGCTGTCTGCTCCCTTACGGGACCCAGGC	1800
QY	1490	-----AGTGGGCACTTGTGCAACCCGGTCACTACTTGTGTGACG	1530
Db	1801	TGACTCTCTTCCCTCTGTAGCTGGCCACCTTGGCAACCCGGTCACTACTTGTGTGACG	1860
QY	1531	CCCGCAGAGATGGGCAACGAGCGCGTCAACTCGCTCGTCACTCCGCGCGCGCACT	1590
Db	1861	CCCGCAGAGATGGGCAACGAGCGCGTCAACTCGCTCGTCACTCCGCGCGCGCACT	1920
QY	1591	GCGGAGGCCAAGACGTCCTTTCTCTC-----	1617
Db	1921	GCGGAGGCCAAGACGTCCTTTCTCTCGTGGCTTCGTGGCAATGATCCGACCGAAT	1980
QY	1618	-----CTTCTGGCCCTGGACCTGTATCTGACGCTTCGAG	1650
Db	1981	AGGCACTACTACCGGACTCTGAGACGTTTCTGCGCTGCACTGTATCTGACGCTTCGAG	2040
QY	1651	GCGCTCGACCTCGCGCGAGTGAAGCTTCAAGAGAGATTGCAACCGCTTCTCCGAG	1710
Db	2041	GCGCTCGACCTCGCGCGAGTGAAGCTTCAAGAGAGATTGCAACCGCTTCTCCGAG	2100
QY	1711	ACTCTCCCTCGAGGACACCTTGGGACTGGGCTCGACGTCAACGCACTTGTGCTCGAGGTC	1770
Db	2101	ACTCTCCCTCGAGGACACCTTGGGACTGGGCTCGACGTCAACGCACTTGTGCTCGAGGTC	2160
QY	1771	AAGAAAGCGCTCAACAAAGCGTCTGAGACAGACAGACGTACGACCTCGAGCGCGAGTGG	1830
Db	2161	AAGAAAGCGCTCAACAAAGCGTCTGAGACAGACAGACGTACGACCTCGAGCGCGAGTGG	2220
QY	1831	CACGACGCTTCTTGTAACGGAACCGGCAACCGTGTGAGCTCTTCTGTCTTCGCTCTCT	1890
Db	2221	CACGACGCTTCTTGTAACGGAACCGGCAACCGTGTGAGCTCTTCTGTCTTCGCTCTCT	2280
QY	1891	GCCAACTCAACCTTACTGCGCGTCAAGCGCTGGAAGGTTGCTCGGCGGAGAGGCAATC	1950
Db	2281	GCCAACTCAACCTTACTGCGCGTCAAGCGCTGGAAGGTTGCTCGGCGGAGAGGCAATC	2340
QY	1951	TGCTTCAACGCGCGAGGTGCGCAACCGCTTGTGACAGACGCGCTTTCGAGGCGCGGAGC	2010
Db	2341	TGCTTCAACGCGCGAGGTGCGCAACCGCTTGTGACAGACGCGCTTTCGAGGCGCGGAGC	2400
QY	2011	CACGCATACCTTTCGCGCGGCAACGCGGCTCTGTAACCTGCTGCGGCGAGAGACTCTCGG	2070
Db	2401	CACGCATACCTTTCGCGCGGCAACGCGGCTCTGTAACCTGCTGCGGCGAGAGACTCTCGG	2460
QY	2071	GTGAGGCGGCGCGGCGCAAGTGTGTTGTGCGGCTGACAGAGAGACGATCCGGAGCAAC	2130

Accession	Sequence	Position
Dp	2461 GTCCAGCCGCGCCGCGCAGCTGTTTTCGCGCTGACGAGGACGATCGGAGCAAC	25220
Qy	2131 GTCTCGGCATCTACGAGGCCATCAAGACGCGCCGATCAACACGCTCTCTGTAAGT	2190
Dp	2521 GTCTCGGCATCTACGAGGCCATCAAGACGCGCCGATCAACACGCTCTCTGTAAGT	2580
Qy	2191 CTGCGCTAAGGCCGAGCAAGCTCTGCTAGACGCGCTCACCCAGAACAAGCTTT	2250
Dp	2581 CTGCGCTAAGGCCGAGCAAGCTCTGCTAGACGCGCTCACCCAGAACAAGCTTT	2640
Qy	2251 CGACGTGCGTGTGTGTCGCAAGACGATCTTCTCTCATACAGATGCGCTTACTCTCG	2310
Dp	2641 CGACGTGCGTGTGTGTCGCAAGACGATCTTCTCTCATACAGATGCGCTTACTCTCG	2700
Qy	2311 CGGTCAACAGTCTCTCAGTCTTTGATGCCGCGTCTT	2351
Dp	2701 CGGTCAACAGTCTCTCAGTCTTTGATGCCGCGTCTT	2741

```

RESULT 5
AAd45824
ID AAd45824 standard; DNA; 2163 BP.
XX
XX
AC
XX AAd45824;
DT
27-DEC-2002 (first entry)
DE Yeast phenylalanine ammonia lyase DNA #2.
XX
XX Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
XX KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
XX KM human cytomegalovirus infection; cytosstatic; virulence; ds.
XX Rhodotorula graminis.
OS
FH FH Location/Qualifiers
FT FT 1. 2163
FT FT *tag= a
FT FT /product= "PAL protein"
FT FT /EC_number= "4.3.1.5"
FT FT /transl_except= (pos:13..15, aa:Xaa)
FT FT /note= "Xaa corresponds to Val, Leu, Phe"
FT FT /transl_except= (pos:34..36, aa:Xaa)
FT FT /note= "Xaa corresponds to Val, Leu, Phe"
FT FT /transl_except= (pos:46..48, aa:Xaa)
FT FT /note= "Xaa corresponds to Val, Leu, Phe"
FT FT /transl_except= (pos:49..51, aa:Xaa)
FT FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT FT /transl_except= (pos:55..57, aa:Xaa)
FT FT /note= "Xaa corresponds to Gly"
FT FT /transl_except= (pos:58..60, aa:Xaa)
FT FT /note= "Xaa corresponds to Ser, Leu or an in-frame stop codon"
FT FT /transl_except= (pos:73..75, aa:Xaa)
FT FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT FT /transl_except= (pos:79..81, aa:Xaa)
FT FT /note= "Xaa corresponds to Pro, Ser"
FT FT /transl_except= (pos:82..84, aa:Xaa)
FT FT /note= "Xaa corresponds to Ala, Pro"
FT FT /transl_except= (pos:100..102, aa:Xaa)
FT FT /note= "Xaa corresponds to Thr"
FT FT /transl_except= (pos:106..108, aa:Xaa)
FT FT /note= "Xaa corresponds to Arg, Ser"
FT FT /transl_except= (pos:115..117, aa:Xaa)
FT FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT FT /transl_except= (pos:118..120, aa:Xaa)
FT FT /note= "Xaa corresponds to Arg, Gly, Trp"
FT FT /transl_except= (pos:142..144, aa:Xaa)
FT FT /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val, Glu, Pro, Leu"
FT FT /transl_except= (pos:160..162, aa:Xaa)
FT FT /note= "Xaa corresponds to Val"

```

```

FT      /transl_except= (pos:166..168, aa:Xaa)
FT      /note= "Xaa corresponds to Lys, Gly, Gln"
FT      /transl_except= (pos:193..195, aa:Xaa)
FT      /note= "Xaa corresponds to Gly, Asp, Val"
FT      /transl_except= (pos:196..198, aa:Xaa)
FT      /note= "Xaa corresponds to Ile, Val, Leu"
FT      /transl_except= (pos:226..228, aa:Xaa)
FT      /note= "Xaa corresponds to Gly"
FT      /transl_except= (pos:259..261, aa:Xaa)
FT      /note= "Xaa corresponds to Thr, Pro, Ser"
FT      /transl_except= (pos:277..279, aa:Xaa)
FT      /note= "Xaa corresponds to Asp, Ala"
FT      /transl_except= (pos:304..306, aa:Xaa)
FT      /note= "Xaa corresponds to Lys, Asn"
FT      /transl_except= (pos:307..309, aa:Xaa)
FT      /note= "Xaa corresponds to Arg, Ser"
FT      /transl_except= (pos:325..327, aa:Xaa)
FT      /note= "Xaa corresponds to Thr, Ala, Ser"
FT      /transl_except= (pos:334..336, aa:Xaa)
FT      /note= "Xaa corresponds to Asp, His, Tyr"
FT      /transl_except= (pos:340..342, aa:Xaa)
FT      /note= "Xaa corresponds to Arg, Ser"
FT      /transl_except= (pos:349..351, aa:Xaa)
FT      /note= "Xaa corresponds to Gly"
FT      /transl_except= (pos:442..444, aa:Xaa)
FT      /note= "Xaa corresponds to Val"
FT      /transl_except= (pos:448..450, aa:Xaa)
FT      /note= "Xaa corresponds to Pro"
FT      /transl_except= (pos:457..459, aa:Xaa)
FT      /note= "Xaa corresponds to Ile, Val, Phe"
FT      /transl_except= (pos:460..462, aa:Xaa)
FT      /note= "Xaa corresponds to Gly, Asp"
FT      /transl_except= (pos:469..471, aa:Xaa)
FT      /note= "Xaa corresponds to Ser, Gly, Arg"
FT      /transl_except= (pos:475..477, aa:Xaa)
FT      /note= "Xaa corresponds to Gly"
FT      /transl_except= (pos:547..549, aa:Xaa)
FT      /note= "Xaa corresponds to Thr"
FT      /transl_except= (pos:667..669, aa:Xaa)
FT      /note= "Xaa corresponds to Pro"
FT      /transl_except= (pos:673..675, aa:Xaa)
FT      /note= "Xaa corresponds to Ser"
FT      /transl_except= (pos:709..711, aa:Xaa)
FT      /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val, Cys, Phe"
FT      /transl_except= (pos:715..717, aa:Xaa)
FT      /note= "Xaa corresponds to Val"
FT      /transl_except= (pos:721..723, aa:Xaa)
FT      /note= "Xaa corresponds to Val"
FT      /transl_except= (pos:724..726, aa:Xaa)
FT      /note= "Xaa corresponds to Val, Leu, Phe"
FT      /transl_except= (pos:736..738, aa:Xaa)
FT      /note= "Xaa corresponds to Lys, Asn, Thr"
FT      /transl_except= (pos:751..753, aa:Xaa)
FT      /note= "Xaa corresponds to Tyr, Ser, Phe"
FT      /transl_except= (pos:775..777, aa:Xaa)
FT      /note= "Xaa corresponds to Leu, Phe"
FT      /transl_except= (pos:913..915, aa:Xaa)
FT      /note= "Xaa corresponds to Leu"
FT      /transl_except= (pos:955..957, aa:Xaa)
FT      /note= "Xaa corresponds to Pro"
FT      /transl_except= (pos:1036..1038, aa:Xaa)
FT      /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp, Ser or a stop codon"
FT      /transl_except= (pos:1231..1233, aa:Xaa)
FT      /note= "Xaa corresponds to Lys, Thr, Met"
FT      /transl_except= (pos:1261..1263, aa:Xaa)
FT      /note= "Xaa corresponds to Thr, Ala, Ser"
FT      /transl_except= (pos:1369..1371, aa:Xaa)
FT      /note= "Xaa corresponds to Leu"
FT      /transl_except= (pos:1372..1374, aa:Xaa)
FT      /note= "Xaa corresponds to Pro"
FT      /transl_except= (pos:1396..1398, aa:Xaa)

```

XX	PS	Claim 7, Page 53-56; 7app; English.
XX	PT	human immunodeficiency virus or human cytomegalovirus infection.
XX	PT	polypeptide, useful for treating a mammal having phenylketonuria, cancer
XX	PT	Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
XX	DR	WPI; 2002-690616/74.
XX	DR	P-PSDB; AAE27943.
XX	PI	Yoshida RK, Kootstra AB;
XX	PA	(PCBU-) PCBU SERVICES INC.
XX	PR	24-JUL-2000; 2000US-00624693.
XX	PR	24-JUL-2001; 2001US-00939408.
XX	PN	US2002102712-A1.
XX	PT	stop codon"
XX	PT	/transl_except= (pos:1939. .1941, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Pro, Ser"
XX	PT	/transl_except= (pos:1948. .1950, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ala, Pro, Ser"
XX	PT	/transl_except= (pos:1978. .1980, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:2122. .2124, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
XX	PT	Tyr, Cys"
XX	PT	/transl_except= (pos:2128. .2130, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ser, Arg, Cys"
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX	PT	/transl_except= (pos:1552. .1554, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr"
XX	PT	/transl_except= (pos:1566. .1608, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Val"
XX	PT	/transl_except= (pos:1666. .1668, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Met, Ile, Val, Leu"
XX	PT	/transl_except= (pos:1669. .1671, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ile, Val, Leu"
XX	PT	/transl_except= (pos:1672. .1674, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Pro"
XX	PT	/transl_except= (pos:1675. .1677, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:1681. .1683, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ile, Leu, Phe"
XX	PT	/transl_except= (pos:1684. .1686, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Lys, Glu, Gln"
XX	PT	/transl_except= (pos:1699. .1701, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:1714. .1716, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Tyr"
XX	PT	/transl_except= (pos:1723. .1725, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Glu, Ala, Val"
XX	PT	/transl_except= (pos:1741. .1743, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:1747. .1749, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Asp, Tyr"
XX	PT	/transl_except= (pos:1849. .1851, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Pro"
XX	PT	/transl_except= (pos:1852. .1854, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Pro, Ser"
XX	PT	/transl_except= (pos:1927. .1929, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX	PT	stop codon"
XX	PT	/transl_except= (pos:1939. .1941, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Pro, Ser"
XX	PT	/transl_except= (pos:1948. .1950, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ala, Pro, Ser"
XX	PT	/transl_except= (pos:1978. .1980, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:2122. .2124, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
XX	PT	Tyr, Cys"
XX	PT	/transl_except= (pos:2128. .2130, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ser, Arg, Cys"
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX	PT	stop codon"
XX	PT	/transl_except= (pos:1939. .1941, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Pro, Ser"
XX	PT	/transl_except= (pos:1948. .1950, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ala, Pro, Ser"
XX	PT	/transl_except= (pos:1978. .1980, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:2122. .2124, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
XX	PT	Tyr, Cys"
XX	PT	/transl_except= (pos:2128. .2130, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ser, Arg, Cys"
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX	PT	stop codon"
XX	PT	/transl_except= (pos:1939. .1941, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Pro, Ser"
XX	PT	/transl_except= (pos:1948. .1950, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ala, Pro, Ser"
XX	PT	/transl_except= (pos:1978. .1980, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:2122. .2124, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
XX	PT	Tyr, Cys"
XX	PT	/transl_except= (pos:2128. .2130, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ser, Arg, Cys"
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX	PT	stop codon"
XX	PT	/transl_except= (pos:1939. .1941, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Pro, Ser"
XX	PT	/transl_except= (pos:1948. .1950, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ala, Pro, Ser"
XX	PT	/transl_except= (pos:1978. .1980, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Thr, Ala, Ser"
XX	PT	/transl_except= (pos:2122. .2124, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
XX	PT	Tyr, Cys"
XX	PT	/transl_except= (pos:2128. .2130, aa:Xaa)
XX	PT	/note= "Xaa corresponds to Ser, Arg, Cys"
XX	PT	/note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
XX		

The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
Cc such proteins. PAL sequences are useful for producing L-phenylalanine.

Query Match 70.4%; Score 1704; DB 6; Length 2163;
Best Local Similarity 84.0%; Pred. No. 3,9e-261;
Matches 1817; Conservative 109; Mismatches 236; Indels 0; Gaps 0;

```
QY 37 ATGGCCCTCTCTTGGAGCTGCTGCGACCAACGCTGCGCAAGGGCTTTACCAAGGGCTCG 96
Db 1 ATGGCCCCCTCCCTCCAGCTGATCGGACCTCGGTGCGCAAGGGCTTTACCAAGGGCTG 60
QY 97 CAGCGGCTCCGACCAAGTGTGCGGCCCCCACTTCGCGCTTCGCGCGACGCGCGG 156
Db 61 CAGCGGCTCCGACCAAGTGTGCGGCCCCCACTTCGCGCTTCGCGCGACGCGCGG 120
QY 157 CTGATGAGCAGCGCGGACCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 216
Db 121 CTGATGAGCAGCGCGGACCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 180
QY 217 CCCACCGACGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 276
Db 181 CCCACCGACGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 240
QY 277 GCGCGCGGACGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 336
Db 241 GCGCGCGGACGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 300
QY 337 GACAGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 396
Db 301 GACAGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 360
QY 397 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 456
Db 361 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 420
QY 457 ATGAGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 516
Db 421 ATGAGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 480
QY 517 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 576
Db 481 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540
QY 577 TGGCTCAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 636
Db 541 TGGCTCAGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 600
QY 637 TTGAACACCGGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 696
Db 601 TTGAACACCGGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 660
QY 697 CTCAGCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 756
Db 661 CTCAGCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
QY 757 GTTTGACAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 816
Db 721 GTTTGACAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 780
QY 817 CTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 876
Db 781 CTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 840
QY 877 TCCGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 936
Db 841 TCCGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 900
QY 937 GCGTTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 996
Db 901 GCGTTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 960
QY 997 ATTCAGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1056
```

```
Db 961 CTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1020
QY 1057 CTCCTTCGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1116
Db 1021 CTCCTTCGAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1080
QY 1117 GAGGAGATTTCTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1176
Db 1081 GAGGAGATTTCTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1140
QY 1177 CTCGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1236
Db 1141 CTCGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1200
QY 1237 GACAAACCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1296
Db 1201 GACAAACCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1260
QY 1297 TCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1356
Db 1261 TCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1320
QY 1357 AACTTCACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1416
Db 1321 AACTTCACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1380
QY 1417 CTGCTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1476
Db 1381 CTGCTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1440
QY 1477 GCTTACGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1536
Db 1441 GCTTACGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1500
QY 1537 GAGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1596
Db 1501 GAGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1560
QY 1597 GCGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1656
Db 1561 GCGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1620
QY 1657 GACTTCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1716
Db 1621 GACTTCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1680
QY 1717 CTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1776
Db 1681 CTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1740
QY 1777 GCGCTCAACAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1836
Db 1741 GCGCTCAACAGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1800
QY 1837 GCGTTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1896
Db 1801 GCGTTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1860
QY 1897 GTCAACCTTACGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1956
Db 1861 GTCAACCTTACGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1920
QY 1957 ACAGCGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2016
Db 1921 ACAGCGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1980
QY 2017 TACCTTCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2076
Db 1981 TACCTTCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2040
QY 2077 GCGCGCGGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2136
```

2041 GCCCGCCGCGGCGAGCTTCTTCTTCGCGAAGAGGTGACGATCGGACCAACGTCTCC 2100
 2137 CGCATCTACGAGGCGCATACAGAGCGCCGATACCACTCTCTGTAAGTCTCGG 2196
 2101 CGCATCTACGAGGCGCATACAGAGCGGCGGATACCACTCTCTGTAAGTCTCGG 2160
 2197 TA 2198
 2161 TA 2162
 RESULT 6
 ABA95244
 ID ABA95244 standard; DNA; 2475 BP.
 ABA95244;
 29-AUG-2003 (revised)
 07-AUG-2003 (revised)
 10-JUN-2002 (first entry)
 Rhodotorula PAL consensus polynucleotide sequence.
 PAL; yeast; phenylalanine ammonia lyase; cinammic acid; cinammate;
 cyostatic; anti-HIV; virocidic; nootropic; dermatological; gene therapy;
 consensus; ds.
 Rhodotorula graminis.
 Rhodotorula mucilaginosa.
 Rhodotorula mucilaginosa.
 Rhodospiridium toruloides.
 Key Location/Qualifiers
 misc_feature 1..2475
 /tag= "a"
 /note= "n= A or C or G or T; "n" indicates no consensus
 at that position"
 W0200208402-A2.
 31-JAN-2002.
 24-JUL-2001; 2001WO-US023270.
 24-JUL-2000; 2000US-00624693.
 (PCBU-) PCBU SERVICES INC.
 Yoshida RK, Koestera AB,
 WPI; 2002-268973/31.
 P-PSDB; ABB07694.
 Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
 treating mammal having disease or disorder from phenylketonuria, cancer,
 human immunodeficiency virus infection and human cytomegalovirus
 infection.
 Claim 5; Fig 2; 135pp; English.
 The invention relates to an isolated and purified yeast phenylalanine
 ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
 host cell comprising the polynucleotide is useful for the production of L
 -PAL, by adding the polynucleotide or the host cell to a composition
 comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
 improving the production of PAL, its analog or another optically active
 unnatural amino acid having PAL-like structure. The PAL polynucleotides
 are useful for treating a mammal having a disease, disorder or condition
 selected from phenylketonuria, cancer, human immunodeficiency virus
 infection and human cytomegalovirus infection. The present sequence
 CC represents a consensus polynucleotide sequence of the Rhodotorula PAL.
 CC (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;
 SQ Query Match 57.3%; Score 1629; DB 6; Length 2475;
 Best Local Similarity 78.0%; Pred. No. 2.9e-249;
 Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
 37 ATGAGCCCTTCTTGAAGTCTGCTGCGACCAAGCTGCGCAACGGCTTTTACCAAGGCTCG 96
 1 ATGAGCCCTTCTTGAAGTCTGCTGCGACCAAGCTGCGCAACGGCTTTTACCAAGGCTCG 60
 97 CAGCGGCTTCGACCAAGTGGGCTGGGGCCCACTTGGGCTTCGGCGGACAGC--CCG 154
 61 CAGCGGCTTCGACCAAGTGGGCTGGGGCCCACTTGGGCTTCGGCGGACAGC--CCG 120
 155 GCGTCGATGGGCAACCGCGGCGCAACGAGTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 214
 121 GCTGCTCTCTCGGACCAACCAAGGAGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 180
 215 ACCCGACCGAC--GACGTCGTCGAGTCTGAGCGGTAACAGCTTACCTGCTGACGTTG 271
 181 ACCCGACCGAC--GACGTCGTCGAGTCTGAGCGGTAACAGCTTACCTGCTGACGTTG 240
 272 TCGGCGCGCGCGCGGACGAGGTCGCGGTC--CAGAAAGAGAGAGATCGCGGCA 330
 241 TCGGCGCGCGCGGACGAGGTCGCGGTC--CAGAAAGAGAGAGATCGCGGCA 300
 331 CGCGTCGACCAAGAGCTGCTGCTTCAAGGCGGCTTCAAGTCTGCTGCTGCTGCTGCTG 390
 301 AAGATTCGACCAAGAGCTGCTGCTTCAAGGCGGCTTCAAGTCTGCTGCTGCTGCTGCTG 360
 391 ACCAGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450
 361 ACAGCTGCTTGGCGGCTGCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 420
 451 GCGCTCATCGAGCAACAGCTCTGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCG 510
 421 GCGCTCATCGAGCAACAGCTCTGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCG 480
 511 GAGCGCGGCTCGAAGACAGCTTCCGCTGAGGTCGTCGCGGTCGCGGTCGCGGTCGCGGTC 570
 481 GAGCGCGGCTCGAAGACAGCTTCCGCTGAGGTCGTCGCGGTCGCGGTCGCGGTCGCGGTC 540
 571 GTCAAGTCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630
 541 GTCAAGTCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 631 AACTCTTGAACCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 690
 601 AACTCTTGAACCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 691 GAGGACCTGAGCGGCTCTGTAATGCGCGGCGCATACCGGTCACCCGAGTCAAG 750
 661 GAGGACCTGAGCGGCTCTGTAATGCGCGGCGCATACCGGTCACCCGAGTCAAG 720
 751 GTTCAAGTTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 810
 721 GTTCAAGTTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
 811 TTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 870
 781 TTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
 871 GCGGTCGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 930
 841 GCGGTCGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
 931 TCGAGGCTTACGAGGCTCTACGAGTGGAGGCAATGATGAGGAGGAGGAGGAGGAGGAGGAG 990
 901 TCGAGGCTTACGAGGCTCTACGAGTGGAGGCAATGATGAGGAGGAGGAGGAGGAGGAGGAG 960
 991 CGTTTATTCAGAGGCTTGGCGCGCGGACCCCGGCAAGGTCAGGTCAGGTCAGGTCAGGTCAG 1050

Db	96	CCNTTCTTCAAGATCAAGCGCCCTCAACCCGACCCAGATGAGAGTGGCCGCAACTC	1020
Qy	1051	CCGAGCTCTTTCCGGCTCGTGTGTTGGCTTTGAGCAAGAGAGATCAAGTCAAG	1110
Db	1021	CGACGCTCTTGGAGGGCAGCANNGTGGCTGCCACACAGAGAGAGTCAAGTCAAG	1080
Qy	1111	GACGACGAGGGCATTTCTTGCCAGGACCGCTACCCGCTCCGAGTGGCTCAGTCTCTC	1170
Db	1081	GACGACGAGGGCATTTCTTGCCAGGACCGCTACCCGCTCCGAGTGGCTCAGTCTCTC	1140
Qy	1171	GGCCGCTCTGTGAGAGACATGATGACGCTTACTGACTCTCTGCTGAGAAC---AAC	1227
Db	1141	GGCCGCTCTGTGAGAGATGATTCACGCCACCGCNGTCTCTGCTGAGAGCGAGAC	1200
Qy	1228	ACGAGACCGGACAACCCGCTCTCGACGTTCAGAAACAAGACACCGGCACACGCGGCAAC	1287
Db	1201	TGACGACCGGACAACCCGCTCTATGACGTTCAGAAACAAGACACCGGCACACGCGGCAAC	1260
Qy	1288	TTCCAGGCGTGGCTGTCTCGATTTTCATGAGAGAAACAGAGCTCGACTCGCCCTCATC	1347
Db	1261	TTCCAGGCGNGCGTGTGCAAAACAGATGAGAGAAATCGCTCTGCTGCGCTGATC	1320
Qy	1348	GGCAAGCTCACTTACGCAATGACAGGATTTGCTCAACGCTGCATGAAACCGCGCTG	1407
Db	1321	GGCAAGCTCACTTACGCAATGACAGGATTTGCTCAACGCGGACATGAAACCGCGCTTN	1380
Qy	1408	CCTTGTCCTCGCTCGCCGAGAACCCGCTCGCTCAACTATCAAGGCAAGGGCTTGGACAT	1467
Db	1381	CCNTCTGCTCTGCTGCGGAGAACCCNTGCTCTCTCTATCACTGAAAGGCTTGACAT	1440
Qy	1468	CACATCGCTTACGCTTGGAGCTGGGACCACTTGCAACCCGCTCACTACCTTGTC	1527
Db	1441	GCCGCGNGCNATCACTTCGAGCTCGAGCACTTGCAACCCGCTTACCAACCAAGTC	1500
Qy	1528	CAGCCCGAGAGATGGGCAACAGGCGGTCAACTCGCTGCTCATCTCCGCGCGCG	1587
Db	1501	CAGCCCGAGAGATGGGCAACAGGCGGTCAACTCGCTGCTCATCTCCGCGCGCG	1560
Qy	1568	ACTGCCAGAGCAAGAGTCTTTCTCTCTCTTTCGCTCGACCTGTACGACGCTC	1647
Db	1561	ACNCGCAGGCGCAAGAGTCTTTCTCTCTCTCTCTCGCACCAACCTCTACTGCTGCTC	1620
Qy	1648	CAGGCGCTGCACTCTCCGCGGATGAGAGTGAATTCATTAAGAAGATTTGACCCGTTCTC	1707
Db	1621	CAGGCGCTGCACTCTCCGCGGATGAGAGTGAATTCATTAAGAAGATTTGACCCGTTNCTC	1680
Qy	1708	CCGACTCTCTCCAGCAGACCTGGGACTGGGCTCGACGTC-----AACGCA	1755
Db	1681	NCNNCGCTCMTCAAGCAGCACTTTGGCNCGCTCGACGCGNNNNNNNNNNNACGAA	1740
Qy	1756	CTTGCGCTCGAGGTCAAGAAAGCGCTCAACAAGGCTCTGAGACAGCAGCAGTACGAC	1815
Db	1741	CTCAGGACAAAGGTCAACAAGNCGCTCMAAAGGACTCAGACAGACCAACTGTACGAC	1800
Qy	1816	CTCAGGCGCGCTGGACGACGCTTTCGTGACGACCGGCAACCGCTCGACGCTC	1875
Db	1801	CTCAGGCGCGCTGGACGACGCTTTCGTGACGACCGGCAACCGCTCGACGCTC	1860
Qy	1876	TGCTCTCGCCCTGTGCA--ACGTCAACCTTACTGCGGTCAACGCGTGAAGTGGCT	1933
Db	1861	NMGTCTCTCGCCNNNNGCCANNAGGTCTGCTCCGCGGCTCAACGCTGGAAGTGGCT	1920
Qy	1934	CGGCGGAGAGGGCATCTGCTCAACGCGGAGATGGCAACCGCTTCTGCAACGCGCT	1993
Db	1921	CGGCGGAGAGGGCATCTGCTCAACGCGGAGATGGCAACCGCTTCTGAGCGCAGT	1980
Qy	1994	CTTGCAAGGCGCGGCGCAACCTCTGCGCGGCAACGCGCTGTGACTCTGTTG	2053
Db	1981	CGTGTGTGCGCCCGCTCNGCTACTCTCGCGCGGCAACGCGCTCTGTACTGTTG	2040
Qy	2054	TGGCGAGAGGCTTGCGCTGCAAGCGCGCGCGCGGCAACGCTGTTTGTGCGCGTCAACAG	2113
Db	2041	TGCGCAGAGGCTTGCGGTCAAGCGCGCGCGCGCGGCAACGCTTCTGCGCAGCAGAGG	2100

QY	2114	AACAGCATGGGAGCAACGCTCGGCATCTACAGAGGCCATCAAGAGCGCCGATCAACC	2179
Db	2101	TGACCATGGGACCAACGCTCTCCGCATCTACAGAGCCATCAAGNCGCGCATCAACC	2166
QY	2174	ACGTCCTGTCAGATGCTCGCGTA-AGGCCGAGCAAGCCTCGCTAGACGCCGCTC	2232
Db	2161	ACGTCCTGTCAGATGCTCGGCTAGGNNCNCNNNNCANNCTCGGNTNNNNCCNNCNC	2220
QY	2233	ACCCGAACACAGCTTTTCGAGCGTCGTGTCGTGCGCAAGACGACTTCCCTCATACA	2292
Db	2221	NNNCNNNNNNNNCTNTTNGNNNTGNNTCATGACNNNNNGAGANNNTNCCNNNNNN	2280
QY	2293	TGTCGCTTACTCTCTTCGCGCTCATCAAGCTCTCAGTCTTTGCTATCCCGCGCTCTC	2352
Db	2281	TNNNNCTNNCTNNCTCNCNNNNANNNGGTGNTNNNNNNCTNNNGNNTNNNNNNC	2340
QY	2353	GGTCGTC-GRACACGTGTATAGAGCCTGGAAATGAGTTGCAAGTCTTGACTTCAAAAA	2411
Db	2341	NGTNNNCANNACNCNTNNNNNNANNCGGANNANGANTNNANGNTTCGGNNC	2400
QY	2412	AAAAAAA 2419	
Db	2401	NNNNAAA 2408	
RESULT 7			
AAD45816			
ID	AAD45816	standard; DNA; 2475 BP.	
XX			
AC	AAD45816;		
XX			
DT	07-AUG-2003	(revised)	
DT	27-DEC-2002	(first entry)	
XX			
DE	Phenylalanine ammonia lyase consensus DNA.		
KM			
KW	Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;		
KM	cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;		
KW	human cytomegalovirus infection; cyostatic; virucide; ds.		
XX			
OS	Rhodotorula graminis.		
OS	Rhodotorula mucilaginosa.		
OS	Rhodospiridium toruloides.		
XX			
PM	US2002102712-A1.		
XX			
PD	01-AUG-2002.		
XX			
PF	24-AUG-2001; 2001US-00939408.		
XX			
PR	24-JUL-2000; 2000US-00624693.		
PR	24-JUL-2001; 2001WO-US023270.		
XX			
FA	(PCBU-) PCBU SERVICES INC.		
XX			
P1	Yoshida RK, Koostra AB;		
XX			
DR	WPI; 2002-690616/74.		
XX			
PT	Novel isolated and purified Rhodotorula phenylalanine ammonia lyase		
PT	polypeptide, useful for treating a mammal having phenylketonuria, cancer,		
PT	human immunodeficiency virus or human cytomegalovirus infection.		
XX			
PS	Claim 5; Page 45-46; 74pp; English.		
XX			
CC	The present invention relates to yeast (e.g. Rhodotorula) phenylalanine		
CC	ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding		
CC	such proteins. PAL sequences are useful for producing L-phenylalanine.		
CC	They are useful for treating mammals having diseases, disorders or		
CC	conditions that would benefit from treatment with PAL proteins such as		
CC	phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or		
CC	human cytomegalovirus (hCMV) infection. Sequences of the invention are		

also used in gene therapy. The present sequence is PAL consensus DNA.
(Updated on 07-AUG-2003 to correct OS field.)

Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

Query Match 67.3%; Score 1629; DB 6; Length 2475;
Best Local Similarity 78.0%; Pred. No. 2,9e-249;
Matches 1678; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

37 ATGGCCCTTCTGGAATCGCTCGCCACCAAGCTCCGCAAGGCTTTTACCAAGGCTCG 96
1 ATGGCCCCCTCAGTCTGATCGGACCTCGGCTCCGCAAGGCTTTTACCAAGGCTCG 60
97 CAGCGCTCTCGACCAAGTCTGCTGGGCTCCGCTTCGCGGCAAGC--CCG 154
61 CAGCGCGGCTCGACCAAGTCTGCTGGGCTCCGCTTCGCGGCAAGC--CCG 120
155 GCGTGAATGGGCAAGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGCTCTCGAGC 214
121 GCTGCTCTCTCCGACCAAGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGCTCTCGAGC 180
215 ACCCGACCGAC--GACGCTGCTGAGCTCAGCGGGTACAGCTTCACCGCTCGTGAAGTTG 271
181 ACCCGACCGAC--GACGCTGCTGAGCTCAGCGGGTACAGCTTCACCGCTCGTGAAGTTG 240
272 TCGGCGCGGCTCGGCAAGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGCTCTCGAGC 330
241 TCGGCGCGGCTCGGCAAGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGCTCTCGAGC 300
331 CCGCTGCAAGAGGCTGCTGCTCTCAAGGCTCCGCTTCGAGAGCTCTCGAGAGTCT 390
301 AAGATGACCAAGAGTCTGCTGCTCTCAAGGCTCCGCTTCGAGAGCTCTCGAGAGTCT 360
391 ACCACGGGCTTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGCTCTCGAGC 450
361 ACGATGCTGCTTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCTCG 420
451 GCGCTCATGACCAAGTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 510
421 GCGCTCATGACCAAGTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 480
511 GAGCGCGGCTTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCTCGC 570
481 GAGCGCGGCTTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCTCGC 540
571 GTCAGTCTGCTCAGCGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 630
541 GTCAGTCTGCTCAGCGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 600
631 AACTTCTGACCAAGGCTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 690
601 AACTTCTGACCAAGGCTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 660
691 GAGCGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 750
661 GAGCGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 720
751 GTTCAAGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 810
721 GTTCAAGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 780
811 TTTGCTCTGAGGCAAGTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 870
781 TTTGCTCTGAGGCAAGTCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 840
871 GCGCTCTGCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 930
841 GCGCTCTGCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 900
931 TCGCAGGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 990
901 TCGCAGGCTTCTGCTGCTGCTGCTCGGCTCGGACCAAGTCTGCTGGGCTTCGAGAGTCT 960

991 CCGTTTATCCAGGCTCTGCGGCTCGGACCCCGGCAAGGCTCGAGGCTCGGCGCAATC 1050
961 CCNTTCTTCAAGGCTCTGCGGCTCGGACCCCGGCAAGGCTCGAGGCTCGGCGCAATC 1020
1051 CGCAGCTCTCTTCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
1021 CGCAGCTCTCTTCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1111 GAGCAGGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
1081 GAGCAGGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1171 GAGCAGGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
1141 GAGCAGGAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
1228 ACGAGACCGCAACCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1287
1201 TCGAGACCGCAACCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
1288 TTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1347
1261 TTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1348 GCGAAGCTCAATTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
1321 GCGAAGCTCAATTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1408 CTTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
1381 CTTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
1468 CACTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
1441 GCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1528 CAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1587
1501 CAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1588 ACTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
1561 ACTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1648 CAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1707
1621 CAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
1708 CCGACTTCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1755
1681 NCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1756 CTTGCGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1815
1741 CTTGCGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
1816 CTTGCGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1875
1801 CTTGCGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1876 TCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1933
1861 NCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1934 CCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1993
1921 CCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1994 CTTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2053
1981 CTTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
2054 TCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2113

2Y 579 GCTGACGCTGGCCACTCGGCGCGCTCGTGCCTTGAAGCGCTCAACCACTTCT 638
 2b 525 CTTAGACCCGCGGCGACCTGGCTGTGCGCTGTGCTTGAAGCGCTTCAACCACTTCT 584
 2Y 639 GAACCAACCGCATCAAGCCCATCGTCCCTCGGCGCTCCATCTCGGCGTGGGCGACT 698
 2b 585 CAACACACGATCAACCCCATCGTCCCTCGGCGGACCATCTCTGCGCGGCGACCT 644
 2Y 699 CAGCCCGCTCTGATCATCGCGCGCGCATCAACCGCTCAACCGGAGCTCAAGTTACGT 758
 2b 645 CTCTCTCTCTCCACATTGACGGCGCATGACGGTCAACCGGAGCAAGGTTGACGT 704
 2Y 759 TTGACACAGAGAAACCGAAGATCATGTTTGCAGCGGAGCCATCTCGCTCTTGGTCT 818
 2b 705 GCTCAACAGAGGCGACAGAGAGATCTGTAGCGCCCGAGAGCGCATGGCGCTTTCAACCT 764
 2Y 819 CGAGGCACTGCTCTCTCGGCGGAGAGAGGTTCTGGTCTGGTCAACGGAACGGCGCTTC 878
 2b 765 CGAGCCCGCTCTCTCGGCGGAGAGAGGTTCTGGTCTGGTCAACGGAACGGCGCTTC 824
 2Y 879 CGGCTCGATGGGAGACCTTCAGTCTGCAAGCTGCAATGCTCTCGCTCTCGAGGCG 938
 2b 825 AGCATCATGAGCCACCTTCGCTCTGCAAGAGCACAATGCTCTCGCTCTCGAGGCT 884
 2Y 939 CTTGACGGCTCTCAAGGTGAGGCGCATGCTGAGCAGAGGCTCTGTCGCGCTTCAAT 998
 2b 885 GCTACGGCGCATGACGGTGAAGGATGTCGAGCGCATCGCGCTCTGTTCCACCCCTTCT 944
 2Y 999 CCAAGACCTCTGCGGCGCGGACCCCGGCGAGTGAAGTGGCGGGAACATTCGCAAGCT 1058
 2b 945 TCACAGCTCAAGCGGCTCAACCGGAGCGAGATGAAGTCCGGGAAACATTCGCAAGCT 1004
 2Y 1059 CTTTCCGCGCTGCTGCTTGGCGTTGAGCAGAGAGAGGTCAAGGTCAAGAGCAGCA 1118
 2b 1005 CTTCAAGAGGAAACCGCTTGTGCTGTCACATGAGAGAGAGGTCAAGGTCAAGAGCAGCA 1064
 2Y 1119 GGGCATTCTTGGCAGGACCGCTACCGGCTCCGCACTGCTCTGATTTCTCGGCGGCT 1178
 2b 1065 GGGCATTCTCGGCGGAGACCGCTACCGCTTGGCACTGCTCTGATGCGTGGCGGCT 1124
 2Y 1179 CGTGAAGACATGATGCAAGCGCTACTGACTCTGCTCGTGAAGAAACA---GACGAC 1235
 2b 1125 CGTCAAGCGACTCATTTCAAGCCGCTGCTTCAACATTCAGAGCGCGGCAAGTGAAGAC 1184
 2Y 1236 CGACAACCGGCTCTGCAAGTGAAGACAGAGACCGGCGAGCGGCAACTTTCAGGC 1295
 2b 1185 CGACAACCGCTCTCATCGAGTGAAGACAGACTTCGACACGCGGCGCAATTTCAGGC 1244
 2Y 1296 GTGCGCTGTCTGATTTGATGAGAGAAACAGAGTGGCACTCGCCCTCATTCGAGCAAGCT 1355
 2b 1245 TCCCGCTGGGCAACACCATGGAAGAACTCGCTCGGCTCGCCCAAGATCGGCAAGCT 1304
 2Y 1356 CAACCTTCAAGAGTGAAGAGTGTCTCAAGCTGCAATGAAACGCGGCGCTGCTTCTGT 1415
 2b 1305 CAACCTTCAAGAGTGAAGAGTGTCTCAAGCTGCAATGAAACGCGGCGCTGCTTCTGT 1364
 2Y 1416 CTTGCTGCGAGGACCGGCTGCTCAACTATCAAGGCAAGGCTTGAACATTCACATCGC 1475
 2b 1365 CTTGCGGCGGAGAGACCGCTGCTCTCTACCACTGCAAGGCGCTTCGACATGCGCGTGC 1424
 2Y 1476 TGGCTTACGCTTGGAGGCTGGGCAACCTTGGCAACCGGCTCACTACCTTCTGTCAGCGCGC 1535
 2b 1425 GGGGTACACTGGAGTTGGAGACCTTGGCAACCTTGGAGAGCATGTTCAGCGCGC 1484
 2Y 1536 AGAGATGGGCAACAGAGCGGTCAACTGCTGCTCACTCTCGGCGCGGCACTGCTGGA 1595
 2b 1485 TGAATGGGGAACAAGGCGGTCACTGCTGCTCACTCTCGGCTCTGCGCAAGACGGA 1544
 2Y 1596 GGGCAACGAGCTCTTCTCTCTCTGCTTGTGCTGCACTGTAAGTCAAGCTTCCAGGCGCT 1655
 2b 1545 GTCAACGAGCTCTTCTCTCTCTGCTGCGCAACCACTTACGCTTCTCAAGCAT 1604

QY 1656 CGACCTTCGGCGAGTAGAGCTCGACTTCAAGAAAGCATTTGACCCGCTTCTCCGACTCT 1715
 Db 1605 CGACTTGCAGCGACAGAGTTGAGTTCAAGAGCAGTTGCGGCCAGGCACTGTCTGCT 1664
 QY 1716 CTTCAACAGACACTCGGCACTGGCCCTGACGTC-----ACGCACTTGCGCT 1763
 Db 1665 CATGACACACACTTGGCTCTGCGCATGACCGGCTGAACTTCGCGAGAGAGTCTGCA 1724
 QY 1764 CGAGTCAAGAAAGCGCTCAACAAAGCTTCGAGCAGACGACGACGATGACCTGAGCTC 1823
 Db 1725 GAAGGTGAACAAGCGTTCGCAAGCGCTCGAGCAGACCAACTGTACGACTGTCTGCC 1784
 QY 1824 GCGTGGCAGACGCTTCTCTGTAACGAGACCGGCAACGCTGTGAGCTCTCTGCTTC 1883
 Db 1785 GCGTGGCAGACGCTTCTCTGTAACGAGACCGGCAACGCTGTGAGCTCTCTGCTTC 1844
 QY 1884 GCGTGGCAGACGCTTCTCTGTAACGAGACCGGCAACGCTGTGAGCTCTCTGCTTC 1943
 Db 1845 GTC-----GCTCTGCTGCGCGCGCGCTCAAGCTTGTGTCGCGCGCTGACTGTC 1895
 QY 1944 GGCATCTGCTCAAGCGGCGAGGTGCGCAACCGCTTCTGCGAAGCGCTCTTCTGCAAGC 2003
 Db 1896 GGCATCTGCTCAAGCGGCGAGGTGCGCAACCGCTTCTGCGAAGCGCTCTTCTGCAAGC 1955
 QY 2004 GCGGCGGAGCATACCTCTGCGCGGCAAGCGGCTCTGCTGCTGCTGCGCGAGGA 2063
 Db 1956 GCGGCGGCTCTGTAACCTTCTGCGCGGCACTCAAGTCTTACGCTTCTGCGCGAGGA 2015
 QY 2064 GCTGCGGCTGAGGCGGCGCGGCGAGCTGTTGTGCGGCTGCAAGAGAGATCGG 2123
 Db 2016 GCTGCGGCTGAGGCGGCGCGGCGAGCTGTTGTGCGGCTGCAAGAGAGATCGG 2075
 QY 2124 GAGCAAGCTCTGCGGCACTTCAAGGCGCATCAAGAGACGCGCGATCAACAGCTCTCGT 2183
 Db 2076 CTCGAAGCTTCAAGATCTTCAAGGCGCATCAAGTGGGAGATCAACAGCTCTCGT 2135
 QY 2184 CAAGTCTGCGGTA 2198
 Db 2136 CAAGTCTGCGTTA 2150

RESULT 9
 AAD33076
 ID AAD33076 standard; DNA; 2151 BP.
 XX
 AC AAD33076;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.
 KW Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TPL;
 KW cytochrome P-450; cytochrome P-450 reductase; liquid crystal polymer;
 KW LCP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst; gene; mutant; variant; ds.
 XX
 OS Rhodotorula glutinis.
 OS Synthetic.
 XX
 FH Key
 FT CDS Location/Qualifiers
 FT 1..2151
 FT /*tag= a
 FT /product= "Rhodotorula glutinis mutant PAL"
 XX
 FN MO200210407-A1.
 XX
 PD 07-FEB-2002.
 XX
 PF 23-JAN-2001; 2001WO-US002099.
 XX
 PR 27-JUL-2000; 2000US-00627216.
 XX

[illegible]

1656 CGACCTCCGCGGATCGAGCTTCAGACTTCAGAGCAGTTCGACCCGCTTCTCCGACTCT 1715

YY		1714	CCTCAGAGACGACTTCGGCACTGGCCCTCGACGTCA-----ACGCATTTGGGCT	1763
YY		1715	GCTTCCAGAGACGACTTCGGCACTGGCCCTCGACGTCA-----ACGCATTTGGGCT	1763
Yb		1665	CATGACACAGCATTTTGSGCTTCGCCAATGAACGGGCTGAACTTGCGCGACGAGCTGTGTA	1724
YY		1764	CGAGGTCAAAGAAGCGCTCTCAACAAGGCTCTTCAGACAGACGACGATACGACTTCGAGACC	1823
Yb		1725	GAAAGTGAACAAGACGCTTCGCGCAAGCGCCCTCGAGCACACCACTGTATGACCTGTGCC	1784
YY		1824	GGCTGAGACAGACCGCTTCTGTACGCGAACCGGACCGGCTGTGAGCTCTCTCGTCTCTC	1883
Yb		1785	GGCTGAGACAGACCGCTTCTGTACGCGAACCGGACCGGCTGTGAGCTCTCTCGTCTCTC	1844
YY		1884	GCCCTCTGCCAAGCTCACCTTACTGTGCGCTCAACGCGTGAAGATTGCTTGCGCGAATA	1943
Yb		1845	GTCTCTGCGCTGCGCGCGCTCAACGCGTGAAGATTGCTTGCGCGAATA	1895
YY		1944	GGCCATCTTCGCTACACGCGCGAGAGTGGCGCAACCGCTTCTGACAGACCGCTCTTCGACAGC	2003
Yb		1896	GGCCATCTTCGCTACACCGCGCGAGAGTGGCGCAACCGCTTCTGAGAGCTCTTCGACAGC	1955
YY		2004	GCCGCGGACAGCATTCCTCTTGCGCGCGACCGCGCTCTGTACTGTTGTGTGGCGAGGA	2063
Yb		1956	GCCCGCGCTCTGTACTCTCTTGCGCGCGACCGCGCTCTGTACTGTTGTGTGGCGAGGA	2015
YY		2064	GCTCGCGCTGACGAGCGCGCGCGCGACGTTGTTGTGTGGCGTGCAGACAGACGATCGG	2123
Yb		2016	GCTTGGCGCTCAAGCGCGCGCGCGCGACGTTGTTGTGTGGCGTGCAGACAGACGATCGG	2075
YY		2124	GAGCAACGCTCTCGCGCATCTACGAGGCGCATTCAGAGACGCGCATACCACTCTCTGT	2183
Yb		2076	CTCGAAGCTCTCAAGATCTACGAGGCGCATTCAGAGACGCGCATACCAAGTACGATCGG	2135
YY		2184	CAAGATGCTCGCGGTA	2198
Yb		2136	CAAGATGCTCGGCTTA	2150
XX	RESULT 10			
XX	AAD33100			
ID	AAD33100	standard; DNA; 2151 BP.		
KX	AAD33100;			
KX	01-JUL-2002	(first entry)		
XZ	R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, Epi8Km-6.			
XX	Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;			
KW	phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;			
KW	cyclochrome p-450; cyclochrome p-450 reductase; liquid crystal polymer;			
KW	ICP; telecommunication; medical device; aerospace application; enzyme;			
KM	biochemical; mutant; variant; gene; ds.			
OS	Rhodotorula glutinis.			
XX	Synthetic.			
XX	Key	Location/Qualifiers		
FH	CDS	1..2151		
FT	/tag= a			
FT	/product= "Rhodotorula glutinis PAL mutant"			
FT	/EC number= "4.3.1.5"			
FT	mutation	replace(645, G)		
FT		/tag= b		
FT	mutation	replace(792, A)		
FT		/tag= c		
FT	mutation	replace(858, T)		
FT		/tag= d		
FT	mutation	replace(1619, T)		
FT		/tag= e		
XX	WO200210407-A1.			
XX				

XX	07-FEB-2002.
XX	23-JAN-2001; 2001MO-US002099.
FR	27-JUL-2000; 2000US-00627216.
XX	(DUPO) DU PONT DE NEMOURS & CO E I.
PA	Gatenby AA, Sariastani FS, Tang X, Qi WW, Vannelli T;
PI	NBI; 2002-280635/32.
DR	P-PDSB; AAE20668.
PT	Newel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT	polyepoxide, or mutant TAL polypeptide, useful for producing
PT	parahydroxycinnamic acid in recombinant host cell lacking cinnamate
XX	hydroxylase.
PS	Disclosure; Page; 139pp. English.
XX	The invention relates to methods for biological production of para-
CC	hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC	bio catalyst for conversion of glucose to PHCA by incorporation of the
CC	wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC	Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC	wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC	to developing a new bio catalyst for conversion of glucose to PHCA by
CC	incorporation of the wildtype PAL which possesses enhanced tyrosine
CC	ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC	plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC	Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC	recombinant techniques. The recombinantly produced PHCA may be used as a
CC	monomer for production of liquid crystal polymers (LCP). LCP may be used
CC	in electronic connectors and telecommunication and aerospace
CC	applications. LCP resistance to sterilizing radiation has also enabled
CC	these materials to be used in medical devices as well as chemical, and
CC	food packaging applications. The present sequence is Rhodotorula glutinis
CC	PAL mutant DNA. Note: This sequence is not shown in the specification,
CC	however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC	(AAB33075) in the sequence listing
XX	Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other:
SQ	
Query Match	49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity	74.0%; Pred. No. 8.6e-181;
Matches 1565; Conservative	1; Mismatches 525; Indels 24; Gaps 3.
QY	99 CGCGCTTCGACCAAGTGGCTGGGGGCCCATCTTGCGCTTCGGCGGACGCGGCGCT 158
DB	45 CGTCGCATCCGAAGAAGCAGCGCTGTCAATGGGCGCTGACCACAACCTTGCAATGCGAGACTTC 104
QY	159 CGATGCGCACGCGCGCGACCAAGTGCAGAGCTGAGATGTGCAGAGACTCTCTCAGCGACC 218
DB	105 GCACCTGCCCAACCAACCAAGTGCAGAGTGCAGATGTGCAGAAAGATGTGTCGCGCGCC 164
QY	219 CACCGACGACGTTGTGTGAGCTTAGCGGGTACAGCTTAACCTTCGTTGAGCTTTCGAGCC 278
DB	165 GACCGACTGACGCTGTGAATCTGACGCGTACTTCCTCAACTCGGAGAGCGTCTTCGCGC 224
QY	279 CGCGCCGAGAGGGCGGAGGCTCCGCTCCACAAGAGAGAGAGATCGCGACGCTGCA 338
DB	225 CGGAGAGAGGGCAGGCGCTGTCCGCTTAAGAGAGAGAGATCCGCTCAAAGATTBA 284
QY	339 CAAGAGCGTGCATTCTCTCAAGGCCACAGCTTCAGAACTCGGTTACGGAATCCACAGGG 398
DB	285 CAATGCGTGAATTTTCGCTCCCACTCTTCATAGAGGTCTACGGCGTCCAGACTGG 344
QY	399 TTTCGTTGCTGGCGCGACACAGAGACTGAGAGATGACATGACGCTCCAGAAAGCGCTCAT 458
DB	345 ATTTGGGAGTATCGCAGAACCCGACCGAAGAGCCATCTCGCTCCAGAAGCTTCTCT 404
QY	459 CGAGACCAAGCTTCTGGCGGTGACCGCGACGCTCGCYCTCGTCTTCAAGCTGGAGCGCG 518

PD 15-FEB-2001.
XX 03-AUG-2000; 2030WO-US021156.
XX 06-AUG-1999; 99US-0147719P.
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX Tang X, Vannelli TM, Qi W, Sarisiani S, Gatenby AA;
XX P-PSDB; AAE16390.
XX WPI: 2002-121549/16.
XX Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT activity.
XX Claim 13; Page 67-68; 75pp; English.
XX The invention relates to a method for producing para-hydroxycinnamic acid
XX (PHCA). The method comprises converting cinnamate to PHCA by converting
XX glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
XX route or by generating a new biocatalyst possessing enhanced tyrosine
XX ammonia-lyase (TAL) activity. The method is useful for the biological
XX production of PHCA which is useful as a monomer for the production of
XX liquid crystal polymers (LCP), where the LCP is used in electronic
XX connectors, telecommunications and aerospace applications. LCP is also
XX useful in medical devices, as well as chemical, and food packing
XX applications due to its resistance to sterilizing radiation. The present
XX sequence is a DNA encoding red yeast, *Rhodotorula glutinis*
XX (*Rhodospiridium toruloides*) mutant PAL derived by substituting Ile at
XX position 540 with Thr. Mutant PAL enzyme possesses enhanced TAL activity
XX
XX Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other:
SQ
Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 8.6e-181;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
QY 99 GCGCGCTCGACCAAGTCGCGGCGGCGCCGACCTCGGCTCGCGCGCAAGCGCGCTC 158
DB 45 CTTCCGATCCGCAAGGCAAGGCTGTATAGGCGCTCGACCACTCGCACTGCGAGCTC 104
QY 159 CGATGCGCAAGCGCGCGCAAGTCGAGCTCGAGATCGTGAGAGAGCTCTTCAAGCAACC 218
DB 105 GCACCTGCCCAACCAACGAGTCACGAGTCGACATCGTCGAGAAAGATGCTCGCGCGCC 164
QY 219 CACCGACAGCTGTCGAGCTTCAAGGCTACAGCTCAAGCTCGGTACGTTGCGCGCG 278
DB 165 GACGACTCGACGCTGAACTGAGGCTGACTCGCTCAACTCGGAAAGCTGCTCGGC 224
QY 279 CCGCGCAAGGCGCGCAGAGGTCGCGCTCAGAAAGCAAGAGATCGCGCAGCGCTCGA 338
DB 225 CCGAGGAAAGGCGAGCGCTGTCGCTCAAGGACAGCGAGATCGCGCTCAAGATGGA 284
QY 339 CAAAGCTCGACTCTCTCAAGGCGCAAGCTTCAAACTCGGTCTTCAAGAGTCAACAGCG 398
DB 285 CAAATCGGTGAGTCTTCTCGCTCGCACTCGCATGAGCGTCTTCAAGCGCTCAAGATGG 344
QY 399 TTTTCGCTGCTGCGCGAGCAGAGAGACTGAGATGACAGCTCGCGCTCGCAAGAGCGCTCAT 458
DB 345 ATTTCGCGATCGAGACACCGGACCGAGAGAGCGCATCTCGCTCAAGAAAGCTTCTCT 404
QY 459 CGAGCAACGACTCTGCGCGGTGACCGCAAGCTCGCTCGCTCTTCAAGCGTGGAGCGCG 518
DB 405 CGAGCAACGACTCTGCGGTGATCTCTCTTCGCTTCAAGTCTCGCTCGCGCTCGCGCG 464
QY 519 CTTTCGAGAACAGCTTCTCGCTTCAAGTGTGTCGCGCGCGCAAGTCACTCGCGCTCAACTC 578
DB 465 TCTTCGAGAACAGCTTCTCGCTTCAAGTGTGTCGCGCGCGCAAGTCACTCGCGCTCAACTC 524
QY 579 GCTTCAGCGCTGAGCACTCGCGCGTTCGCGCTCGCTTCAAGGCGCTCAACCACTTCTT 638

DB 525 CTTGACCCCGGCGCACTCGGCTGTCCGCGCTCGCTGCTCGAGGCGCTCAACCACTTCT 584
QY 639 GAACCAACCGCATCAGCGCATCTGTTCCCTTCGCGGCTCCATCTCGCGCTCGGCGCACT 698
DB 585 CAACCAAGGATCACCCCGCATCTGTTCCCTTCGCGGCTCCATCTCGCGCTCGGCGCACT 644
QY 699 CAGCGCGCTCTGACATCGCGCGGCGCATCAAGGCTCAAGCGCGCGCAAGGTTGACGT 758
DB 645 CTTCTCTCTCTCTCACTTCAAGGCGCATCAAGGCTCAAGCGCGCGCAAGGTTGACGT 704
QY 759 TTTGACAGAGGAAACCGAAGATCATGTTTGGCGCGGAGCGCATCTCGCTTGTGCTCT 818
DB 705 CGTCCACAGAGGCAAGAGAAAGATCTGTACGCGCGGAGCGCATGCGCTCTTCAACT 764
QY 819 CGAGGAGCTGCTCTCGCGCGCGGAGGAGGCTCGGTCTGTCAGAGGAGCGCGCTCTC 878
DB 765 CGAGCGGCTGCTCTCTCGCGCGGAGGAGGAGGCTGCTGTCAGAGGAGCGCGCTCTC 824
QY 879 CGCTCGATGCGCAACCTCAATGTCAGCACTCGCAATGCTCTGCTCTCTCGAGGC 938
DB 825 AGCATGATGCGCAACCTCGCTGTCAGCACTCGCAATGCTCTGCTCTCTCGAGGC 884
QY 939 CTTGAGGCTCTCAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 998
DB 885 GCTCAGCGCATGACGATGAGAGGATGAGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 944
QY 999 CCAGAGCTGTCGCGCGCGCAACCGCGCGAGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1058
DB 945 TCAGAGCTCAGCGCGCTCTCAACCGCGCGAGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1004
QY 1059 CTTTCGCGCTGCTGTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 1118
DB 1005 CTTTCGCGCGAGAGCGCGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 1064
QY 1119 GGGCATTTTGGCGAGAGCGCTTACCGCGCTCGCGAGCTCGCTTCTCGAGCGCT 1178
DB 1065 GGGCATTTTGGCGAGAGCGCTTACCGCGCTCGCGAGCTCGCTTCTCGAGCGCT 1124
QY 1179 CTTGAGGACATGATGACGCGCTTCTGACTTCTGCTCGAGAACAA--CGAGAC 1235
DB 1125 CTTGAGGACATGATGACGCGCTTCTGACTTCTGCTCGAGAACAA--CGAGAC 1184
QY 1236 CGACACCGCGCTCTCTGACGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1295
DB 1185 CGACACCGCGCTCTCTGACGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1244
QY 1236 GTGCGCTGCTCGATTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
DB 1245 TGCCGCTGTCGCAACACATGAG 1304
QY 1356 CAATTCAG 1415
DB 1305 CAATTCAG 1364
QY 1416 CTTTCGCGCGAG 1475
DB 1365 CTTTCGCGCGAG 1424
QY 1476 TGCTTACGCTTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 1535
DB 1425 GGGCTACACCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 1484
QY 1536 AGAGATGAGCAACAG 1595
DB 1485 TGAGATGAGCAACAG 1544
QY 1596 GGGCAACAGAGTCTTCT 1655
DB 1545 GTTCAGAGAGAGTCTTCT 1604
QY 1656 CGACTTCGCGCGAGAGAGTTCGACTTCAAGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTC 1715

Db 645 CTCTCTCTCTCTCTCAATTGCAAGGCGCATCAGCGCTCACCCGGAACAGAAAGTGACGCT 704
2Y 759 TTGGACGAGGAAACGGAAGATCATGTTTGGCCGAGAGCCATCTCGTCTTTGATCT 818
Db 705 CGTCACGAGGAAAGGAGAGATCTGTACGCCGGAGGCAATGGCGCTTTCAACCT 764
2Y 819 CGAGGCACTGCTCTCGGCGCCGGAAGAGGCTGTGATGATCAACGGAACGCGCGCTTC 878
Db 765 CGAGCCCTGCTCTCTCGGCGCCGGAAGAGGCTGTGATGATCAACGGAACGCGCGCTTC 824
2Y 879 CGCTCGATGAGCGAACCCTCACTGTGCAAGACTCGCAATGCTCTCGTCTCTTCGAGGC 938
Db 825 AGCATCGATGAGCGAACCCTCGCTGTGCAAGAGCATGCTCTCGTCTCTTCGAGGC 884
2Y 939 CTGAGCGGCTCTCAAGGCTGAGGAGCATGATGAGCGAGGCGCTCTGTCGCGCTTCAT 998
Db 885 GCTCAGGCGCTCAAGGCTGAGGAGCATGATGAGCGAGGCGCTCTGTCGCGCTTCAT 944
2Y 999 CCACGACGCTGCGCGCGCCGCAACCCGCGCAGGTGAGGTGCGCGCAACATCCGACGCT 1058
Db 945 TCAGACGCTCAGCGCGCCCTCAACCCGACGAGATCGAAGTGGCGGAAACATCCGCAAGCT 1004
2Y 1059 CCTTCCGCGCTGTGCTGTTGGCTGTGAGCAAGAGAGAGTCAAGGTCAGAGACGACGA 1118
Db 1005 CCTCAGAGGAAAGCGGCTTGTGTCAACATGAGAGAGAGTCAAGGTCAGAGACGACGA 1064
2Y 1119 GGGCATCTTCTGCGAGAGACCGCTACCGGCTCGACGCTGCGCTCACTTCTCGCGCGCT 1178
Db 1065 GGGCATCTTCTGCGAGAGACCGCTACCGGCTCGACGCTGCGCTCACTTCTCGCGCGCT 1124
2Y 1179 CGTGAGAGCATGATGACGCGCTTACTGACTCTCTCGTCAAGAACAA---GACGAC 1235
Db 1125 CGTCAGGACCATCTTCAAGCGCCAGCGCTCTCAACATCGAGGCGCGGACGTCAGAC 1184
2Y 1236 CGACAGCGCGCTCTCGAGCGCTGAGAAACAGACGCGCGAGCGGCAACTTCCAGGC 1295
Db 1185 CGACAGCGCTCTCTCATGACGCTGAGAAACAGACTTCCAGCAGCGCGCAATTTCCAGGC 1244
2Y 1296 GTGCGCTGTCTGATTTTGATGAGAGAACAGGCTGCACTCGCGCTCATCGGCAAGCT 1355
Db 1245 TGCGCTGTGGCCAAACCATGAGAAAGACTCGCTCGGCTCGCCAGATCGGCAAGCT 1304
2Y 1356 CACTTTCAGAGTGCAGACCGAGTGTCTAAGCTGCAATGAAACGCGCGCTCGCTGTG 1415
Db 1305 CACTTTCAGAGTGCAGACCGAGTGTCTAAGCTGCAATGAAACGCGCGCTCGCTGTG 1364
2Y 1416 CTTGCTGCGAGAGACCGGCTGCTCACTATCAGGCAAGGCTTGGACATTCATCATGCG 1475
Db 1365 CTTGCTGCGAGAGACCGGCTGCTCTCTAACAATGAAAGGCTTCGACATCGCGCTGC 1424
2Y 1476 TGCTTAAGCTTGGAGCTGGCGCACTTTCGCAACCGGTCATCTTGTCCAGCGCGC 1535
Db 1425 GCGGTACACCTCGGAGTTGGGACACTTGCCAAACCTGTGTGACAGCATGTCCAGCGGC 1484
2Y 1536 AAGAGATGGGCAACAGAGCGGTCAACTCGCTGCTCTATCTCCGCGCGCGCATGTCGA 1595
Db 1485 TGAAGATGGGCAACAGAGCGGTCAACTCGCTGCTCTATCTCCGCGCGCGCATGTCGA 1544
2Y 1596 GGCAGACGAGCTCTTCTCTCTCTCTGCTCGACCTGTATGCAAGCTTCCAGGCGCT 1655
Db 1545 GTTCAACGAGCTCTTCTCTCTCTCTCTCGCAACCACTTACTAGCTTCCAAAGCAT 1604
2Y 1656 CGACCTCCGCGAGTGGAGCTGACTTCAAGAGAGTTCGACCGGCTTCCCGCACTCT 1715
Db 1605 CGACTTGGCGGACCGAGTTGAGTTCAAGAGAGTTGGCGCGCGCATGTCTGCT 1664
2Y 1716 CTTCCAGCAGCACTTGGCACTGGCTCGAGTCA-----ACGCACTTGGCT 1763
Db 1665 CATCGACGAGCACTTGGCTCGCATGACGCGGCTGAAACCTGGCGCAGAGCTGTGCA 1724
2Y 1764 CGAGGTCAAGAGGCGCTCAACAGAGCTTCAAGAGAGCGAGAGCTTCAAGCTTCAAGCC 1823
Db 1725 GAAGGTGAACAAGAGCTCGCAAGCGCTCGAGAGAGCAAACTGTGATGACGCTGTGCC 1784

QY 1824 GCGGTGACAGACGCTTCTGTACGAGACCGGCAACGCTGTGAGACTCTCTGCTCTC 1883
Db 1785 GCGGTGACAGACGCTTCTGTACGAGACCGGCAACGCTGTGAGACTCTCTGCTCTC 1844
QY 1884 GCGCTTCCCAACGCTCACTCTTACTGCTGCTCAACGCTGGAAGCTTCTCGCGCAAGA 1943
Db 1845 GTC-----GCTCTGCTCGCGCGCTCAACGCTGGAAGCTGCGCGCGCTCAAGTC 1895
QY 1944 GGCATCTGCTCAACGCGAGAGGTCGCAACGCTTCTGACAGACCGCTTCTCGAGGC 2003
Db 1896 GGCATCTGCTCAACGCGAGAGGTCGCAACGCTTCTGACAGACCGCTTCTCGAGGC 1955
QY 2004 GCGCGCGCAGCATACCTCTGCGCGCGCAGCGGCTCTGTACTCTGTGTCGCGAGGA 2063
Db 1956 GCGCGCGCTCTGACTCTCTGCGCGCAGCTCAGATCTCTTACGCTTCTGTCGCGAGGA 2015
QY 2064 GCTGCGGTGAGAGCGCGCGCGCGCGAGCTTGTGTCGCGGTGTCAGAGAGAGATCGG 2123
Db 2016 GCTTGGGTCAAGGCGCGCGCGAGAGCTTCTCTGCGCAAGAGAGTGAAGATCGG 2075
QY 2124 GAGCAAGCTCTCGGCGCATCTACGAGGCGCATCAAGAGAGCGCGCATCAACGCTCTGCT 2183
Db 2076 CTGCAAGCTCTCAAGATCTTACGAGGCGCATCAAGTGGGCGAGATCAACAGCTCTCTCT 2135
QY 2184 CAAGATGCTCGCTA 2198
Db 2136 CAAGATGCTCGCTA 2150

RESULT 13
ADE36055
ID ADE36055 standard; DNA; 2151 BP.
XX
XX ADE36055;
XX
XX
XX 29-JAN-2004 (first entry)
XX
XX Rhodospiridium glutinis EPr8km-6 mutant PAL encoding DNA SEQ ID NO:3.
XX
XX recombination; nucleic acid template; recombined polypeptide;
XX
XX altered property; combined gene; gene; ds.
XX
XX Rhodospiridium.
XX
XX WO2003072743-A2.
XX
XX
XX 04-SEP-2003.
XX
XX 26-FEB-2003; 2003WO-US005708.
XX
XX 26-FEB-2002; 2002US-0360279P.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Milano J, Tang X;
XX
XX PI
XX
XX DR MPI: 2003-721766/68.
XX
XX DR P-PSDB; ADE36056.
XX
XX
XX Recombination of nucleic acid templates, useful e.g. for producing
XX PT enzymes with altered properties, is based on cyclic extension of unpaired
XX PT primers.
XX
XX
XX PS Example 2; SEQ ID NO 3; 481bp; English.
XX
XX
XX The present invention describes a method for the recombination of nucleic
XX CC acid (NA) templates. The method comprises: (a) providing at least two
XX CC double-stranded templates (T1, T2) having different 5' and 3' regions in
XX CC their sense strands; (b) contacting the templates with at least one each
XX CC of forward and reverse primers that anneal, respectively, only to the 3'-
XX CC region of the antisense strand of T1 and to the 3'-region of the sense
XX CC strand of T2; (c) extending the primers by no more than 1000 nucleotides

CC (nt): (d) separating extended primers from their templates; (e)
CC reannealing the extended primers, to either template; and (f) repeating
CC cycles of steps (b)-(e) until at least one full-length extension product,
CC i.e. a recombinant of T1 and T2, is obtained. Also described: (1) a
CC method starting from at least one each of antisense and sense single-
CC stranded templates; (2) generating a recombinant polypeptide (I) with
CC altered properties by expressing a recombinant polypeptide (I) with
CC processes and screening the polypeptides formed for properties different
CC from those of polypeptides from either of the templates; and (3) (1)
CC produced by method (2). The methods are useful for the recombination of
CC nucleic acid templates of interest, and generation of a recombinant
CC polypeptide having altered properties. The method can be used to make
CC combined genes that express proteins e.g. enzymes, cytokines, growth
CC factors, viral proteins or microbial antigens with altered properties.
CC e.g. stability, activity or specificity. The present sequence is used in
CC the exemplification of the present invention.

XX Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 49.5%; Score 1196.6; DB 9; Length 2151;

Best Local Similarity 74.0%; Pred. No. 8.6e-181; Mismatches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCGACCACTGCTGCGGCGCCCACTTGGCTCTCGCGCGACGCCGCGCT 158
DB 45 CGTCGATCCGCAAGCAGGCTGTCAATGGCGCTCGACCACTCGAGTGGAGGCTC 104
QY 159 CGATGGCCAGCGCGCGACCACTCGGACTCGAGATCTGACAGAGCTCTCAAGCAC 218
DB 105 GCACTGCGCCACCAACCCAGGTACCGAGGTGACATCGTCAGAAAGTGTCTGCGCGCC 164
QY 219 CACCGACGCTGCTGAGTCAAGGCTGACAGCTCAAGCTCCGCTGACGCTGCGCGC 278
DB 165 GACCGACTGAGAGCTGAACTGAGGCTACTCGGTCACTCGGAAACGTCGCTGCGC 224
QY 279 CGCGCGCAAGGCGCGAGGCTGCGGTCCAGAGACGACGAGATCCGCGACGCTGCA 338
DB 225 CGGAGGAAGGCGAGGCTGTCTCGGTCAAGAGACGAGATCCGCTCAAGAATTGA 284
QY 339 CAAGAGCTGCACTCTCTCAAGGCGGAGCTTCAGAACTCGGTCAAGAGTCAACGAG 398
DB 285 CAATTCGTGAGTCTTGTGCGCTGCACTCTCAATGAGGCTCTACGCGCTCAAGACTG 344
QY 399 TTTGGTGTGCGCGGACAGAGACTGAGATGCAATCAAGCTCCAGAAAGCGCTCAT 458
DB 345 ATTGGCGGATCCGAGACACCGGACCGAGAGACGCAATCGCTCAGAAAGCTCTCT 404
QY 459 CGAGCAGCAGCTCTGCGGCGTGAAGCGCGACGTCGCTCGCTTCGAGTCCGAGAGCG 518
DB 405 CGAGCAGCAGCTCTGCGGCTGCTCTCTCTGCTGAGCTCGTTCGCTCGAGCGCG 464
QY 519 CCTGAGAACAGCTTCGCTGAGTGTGCGGCGGACATGATGTCATCCGCTCAATC 578
DB 465 TCTGAGAACAGCTTCGCTGAGTGTGCGGCGGACATGATGTCATCCGCTCAACG 524
QY 579 GGTCAAGGCGGACCTGCGGCGCTGCGCTGCTGCTGAGGCGCTCACTTCTT 638
DB 525 CTGACCGCGGCGCTGCGGCTGCTGCTGCTGCTGAGGCGCTCACTTCTTCT 584
QY 639 GAACACGCGATCAAGCCATGCTCCGCTCGGCGCTCATCTGCGCTCGAGCGACT 698
DB 585 CAACACGCGATCAAGCCATGCTCCGCTCGGCGGACATCTCTGCGCTCGAGCGACT 644
QY 699 CAGCGCGCTCTGTCATGCGCGGCGGACATCAAGGTCAACCGGCTCAAGTTCAGCT 758
DB 645 CTCTCTCTCTCTCACTTTCAGGCGGCGATCAAGGCTCAACCGGCTCAAGTTCAGCT 704
QY 759 TTTCAGAGGAGAACGAGATCATGTTGCGGCGGAGGCGATCGCTCTTGGTCT 818
DB 705 CGTTCACAGAGGAGAGAGATCTGTAGCGCGCGAGGCGATGCGCTCTTCAACT 764
QY 819 CGAGCAGTCTCTCTGCGCGGAGAGGCTCTGCTGATCAACGAGACGCGCTCTC 878

DB 765 CGAGCGCGCTCTGCTCGGCGGAGAGGCTCTGCTGATCAACGAGCGCGCTCTC 824
QY 879 CGCCTGAGAGGAGACCTCACTGTCAGCACTGCAATGCTCTGCTCTCGAGGCG 938
DB 825 ACATCATGAGGAGACCTCTGCTGTCAGAGAGCAATGCTCTGCTCTCTCGAGTC 884
QY 939 CTTCAGGCTCTCAAGGTGAGGAGCTGTCGCGCAGACAGGCTCTGCGCGCTTCAT 998
DB 885 GCTCAGCGCATGACGCTGAGAGGATGTCGAGCAGCGCGCTCTGTCACCCCTTCT 944
QY 999 CCAGAGCTGCGCGCGCGCGCGCGCGCGCGAGGTGAGGTGCGCGGCAATTCGAGCGCT 1058
DB 945 TCAGACGTCACGCGCGCTCAACCGAGCATGATGAGTTCGCGGAACTCCGAACT 1004
QY 1059 CTTTCCGCGCTGCTGCTTGGCTTGAAGCAAGAGAGAGTCAAGTCAAGAGCAGCA 1118
DB 1005 CTTGAGGAGAGCGCTTGTGCTGTCACCATGAGAGAGTCAAGTCAAGAGCAGCA 1064
QY 1119 GGGCATTTTGGCAGAGACCGCTACCGCTCGGACAGTCCCTCACTTCTGCGCGCT 1178
DB 1065 GGGCATTTTGGCAGAGACCGCTACCGCTCGGACAGTCCCTCACTTCTGCGCGCT 1124
QY 1179 CGTGAGAGCATGATGACGCTTACTGACTCTCTGCTGAGAAACA---CGACGAC 1235
DB 1125 CGTCAGGACCTCATTCAGCGCGCACCGCGCTCTCACCATGAGGCGGCGCATGACGAC 1184
QY 1236 CGAACACCGCTCTCTGACGCTGAGAAACAGACGACCGCGCATCTTCCAGCG 1295
DB 1185 CGAACACCTTCTTCAGCGCTGAGAAACAGACTTTCAGCACCGCGGCGCAATTTCCAGCG 1244
QY 1296 GTGCGCTGTCTGATTCGATGAGAGAGACAGAGCTCGACCTGCGCTCATGCGAGCT 1355
DB 1245 TGCGGCTGTGCGCAACACATGAGAGAGCTGCTGCGGCTGCGGCAATGCGAGCT 1304
QY 1356 CACTTACGAGTGCACCGAGTTGCTCAACGCTGCTCATGAACCGCGGCTGCTTCTG 1415
DB 1305 CACTTACGAGTGCACCGAGTTGCTCAACCGCTGCTCATGAACCGCGGCTGCTTCTG 1364
QY 1416 CTTGCTGCGAGGAGACCGGCTGCTCACTTCAACCGGCAAGGCTTGAATTCATTCATGCG 1475
DB 1365 CTTGCTGCGAGGAGACCGGCTGCTCTCTCTCACTTCAACCGGCTGCTGAGATGCGGCTG 1424
QY 1476 TGCTTACGCTTTCGAGCTGCGGCACTTTCGCAACCGGCTCATTACTTCTGTCAGCGCG 1535
DB 1425 GCGGTACACCTCGAGTTGGAGACCTCGGCAACCTGAGAGAGCAGATGTCAGCGCG 1484
QY 1536 AGAATGAGGACACGAGCGCTCACTTCTGCTGCTCTCATCTCGGCGCGGACATGCGCA 1595
DB 1485 TGAATGAGGACACGAGCGGTCACTGCTGCTGCTCATCTGCTGCTGCGGACAGCGCA 1544
QY 1596 GAGCAGAGGCTCTTCTCTCTCTCTGCTTCTGCTGCACTGATCTGACGCTCAAGCGCT 1655
DB 1545 GTCCAGAGAGTCTTCTCTCTCTCTCTGCTGCAACCGCTCTGATGCTTCTCAAGCAT 1604
QY 1656 GAGCTCCGCGGATGAGTCACTTCAAGAGAGCTTCAACCGGCTTCTCCAGCTCT 1715
DB 1605 GAGCTTCCGCGGACAGAGTTCAGATTCAAGAGAGTTCGCGCCAGCATGCTGCT 1664
QY 1716 CTTGAGAGACCTCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1763
DB 1665 CATGAGACACATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
QY 1764 CGAGTCAAGAGCGCTCAACAGGCTTCAGAGACGAGACGAGATGACGCTTCAAGCT 1823
DB 1725 GAGGTGAGAGAGCGCTTCGCAAGGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1784
QY 1824 GCGGTGAGAGAGAGCGCTTCGCAAGGCTTCGCAAGGCTTCGCAAGGCTTCGCAAGG 1883
DB 1785 GCGGTGAGAGAGAGCGCTTCGCAAGGCTTCGCAAGGCTTCGCAAGGCTTCGCAAGG 1844
QY 1884 GCGCTTTCGAG 1943
DB 1845 GTC-----GCTCTGCTGCGCGCGCTGCAACGCTGAGAGAGTTCGCGCGCGAGAGTC 1895

Query Match	Score	DB 1	Length
	49.58		2331

1119 GGGCATCTTCGCCAGGACCGCTACCCGCTCCGCACGTGGCCTCAGTTCCTGGCCCGCT 117

Db 165 GACCGACTCGACGCTCGAATCGACGCGCTACTGCTCAACCTCGAGAGAGCTGCTCGGC 224
 2y CGCCCGGAGGGGGGCGAGGCGCGCGCTCCAGAACGAGAGATCCGGACGCGCTCGA 338
 Db 219 CGCAGAGAAAGGCGAGCGCTGCTCCGCTCAAGAGACGAGAGATCCGCTCAAAAGTTAA 284
 2y 339 CAAGAGCTCGACTCTCTCAAGGCGCGACTTCAGAACTCGGTCTACGAGATCAACGCG 398
 Db 285 CAATTCGGTCAAGTTCTTGGCGCTCGCAACTCTCATAGAGGTCTACGGGCTCAAGCTGG 344
 2y 399 TTTGGTGGCTCGGCGCAGAGAGAGATGAGATGCAAGTCAAGCTCCAGAAAGCGCTCAT 458
 Db 345 ATTGGCGAATCCGACAGACCGCCACCGAGAGAGAGCATCTCGCTCAAGAGCTCTCT 404
 2y 459 CGAGACGAGCTTCGCGCGCTGACGCGGACGCTCGCTCTCTTCAAGCTTCGAGCGCG 518
 Db 405 CGAGACGAGCTTCGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
 2y 519 CCGTGAAGAACGCTTCGCTCGAGAGTGTCCGCGCGCGCATGTGATCCGCGTCAACTC 578
 Db 465 TCTGAGAACTCGCTTCCCTCGAGGTTGTTGCGCGCGCATGCAATCCGCGTCAACAG 524
 2y 579 GCTCAGCGGCGCGCATCGGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 638
 Db 535 CTTGACCGCGCGCATCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
 2y 639 GAACACGCGCATCAGCGCCATCGTCCCTCGCGCGCTCGATCTCGGCGTGGCGACCT 698
 Db 585 CAACACGCGCATCAGCGCCATCGTCCCTCGCGCGCTCGATCTCGGCGTGGCGACCT 644
 2y 699 CAGCCCGCTCTGTACATCGCGCGCGCATCAGCGGTCAACCCGAGCATCAAGGTTACGT 758
 Db 645 CT 704
 2y 759 TTTGACGAGGAGACCGAAGATCATGTTGCGCGCGAGGCGCATCTCGCTCTTGGCT 818
 Db 705 CCGTCAAGAGGAGAGAGAGATCTGTACCGCCGCGAGGCGATGGCGCTTCAACCT 764
 2y 819 CGAGGAGTGTCTCTCGCGCGCGAGAGGCTGTGCTGTCTCAACGAGCGCGCTCTC 878
 Db 765 CGAGCGCGCTCGCTCGCGCGCGAGAGGCTGTGCTGTCTCAACGAGCGCGCGCTCTC 824
 2y 879 CGCGTGTAGTGGGAGCGCTCTGATCTGACAGCTCGAGCATGCTCGCTCGCTCGAGG 938
 Db 825 AGATGAGTGGGAGCGCGCTGTCTGTGACAGCAGCATGCTCTGCTCTCTCTCTCTCT 884
 2y 939 CTGAGCGCTCTCAGCGTGAAGGCGATGCTCGCGCGAGAGGCGTGTGCGCGCTCAT 998
 Db 885 GCTCAGGCGCATGAGCGTGAAGCGATGCTGTGCGACGCGCGCTGTCTCAACCTTCT 944
 2y 999 CCAAGAGTGTGCG 1058
 Db 945 TCAAGAGTGTGCG 1004
 2y 1059 CTTTTCCGCGTGTGCTTGGCGTGTGAGCAGAGAGAGAGAGTCAAGTCAAGAGAGAG 1118
 Db 1005 CTTGAGAGAGAGAGCGCGCTTGTCTGTCAATGAGAGAGAGTCAAGAGAGAGAG 1064
 2y 1119 GGGCATTTCTTCCAGAGACCGCTACCGCGTCCGAGCGTCTCAAGTCTCTCGCGCGCT 1178
 Db 1065 GGGGATTTCTTCCAGAGACCGCTACCGCGTCCGAGCGTCTCAAGTCTCTCGCGCGCT 1124
 2y 1179 CTTGAGAGAGATATGAGAGCGCTTCTGAGCTGTCTGCGTCAAGAGAGAGAGAGAG 1235
 Db 1125 CTTGAGAGAGATATGAGAGCGCTTCTGAGCTGTCTGCGTCAAGAGAGAGAGAGAG 1184
 2y 1236 CGAAGAGCGCGCTCTCGAGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295
 Db 1185 CGAGAGAGCGCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
 2y 1296 GTGCGTGTCTGATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
 Db 1245 TGCGGCTGTGGCCAAACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304

Qy 1356 CAATTGAG 1415
 Db 1305 CAATTGAG 1364
 Qy 1416 CTTGAG 1475
 Db 1365 CTTGAG 1424
 Qy 1476 TGTGAG 1535
 Db 1425 GCGTCAACCTCGAG 1484
 Qy 1536 AGAATGAG 1595
 Db 1485 TGAATGAG 1544
 Qy 1596 GAGCAAG 1655
 Db 1545 GTCCAG 1604
 Qy 1656 GAGCTCGCGCGAG 1715
 Db 1605 GAGCTCGCGCGAG 1664
 Qy 1716 CCGTCAAG 1763
 Db 1665 CATGAG 1724
 Qy 1764 CGAGTCAAG 1823
 Db 1725 GAGGAG 1784
 Qy 1824 GAGGAG 1883
 Db 1785 GAGGAG 1844
 Qy 1884 GCGCTCTGCAAG 1943
 Db 1845 GTC-----GCTCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1895
 Qy 1944 GCGCATCTGCTCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2003
 Db 1896 GCGCATCTGCTCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1955
 Qy 2004 GCGCGAG 2063
 Db 1956 GCGCGAG 2015
 Qy 2064 GCTGCGAG 2123
 Db 2016 GCTGCGAG 2075
 Qy 2124 GAGCAAG 2183
 Db 2076 CTGCAAG 2135
 Qy 2184 CAAGATGCTCGGTA 2198
 Db 2136 CAAGATGCTCGGTA 2150

Search completed: September 11, 2004, 14:50:22
 Job time : 1351.06 secs

Db 301 CGCGTCCAGAACGACGAGATCCGCGACGCGCTGACACAGAGCGTGCATCTTCCCTCAAG 360
QY GCCAGCTTCAGAACTCGGTCTACGAGAGTACACAGGGTTTGGGTGGCTCGGCGAGACG 420
Db 361 GCCAGCTTCAGAACTCGGTCTACGAGAGTACACAGGGTTTGGGTGGCTCGGCGAGACG 420
QY AGAAGCTGAGAGTACAGTACAGCTTCAGAAAGCGCTTCATGAGCACACAGCTTCGCGCGTG 480
Db 421 AGAAGCTGAGAGTACAGTACAGCTTCAGAAAGCGCTTCATGAGCACACAGCTTCGCGCGTG 480
QY ACCGCGACGTCCGCTCGCTTACAGGTCCGAGCGCGCTTCGAGAAACAGCTTCGCGTG 540
Db 481 ACCGCGACGTCCGCTCGCTTACAGGTCCGAGCGCGCTTCGAGAAACAGCTTCGCGTG 540
QY GAGGTGCTCCGCGCGCGCATGTGATCCGCTCAACTCGCTCACCGCTGCGCATCGCGC 600
Db 541 GAGGTGCTCCGCGCGCGCATGTGATCCGCTCAACTCGCTCACCGCTGCGCATCGCGC 600
QY GTCGCGCTCGTCTCTTGAAGCGCTCACCACTTCTTGAACACCGCATACCGCCATC 660
Db 601 GTCGCGCTCGTCTCTTGAAGCGCTCACCACTTCTTGAACACCGCATACCGCCATC 660
QY GTCCCGCTCGCGGCTCCATCTCGGCGTCCGCGGACCTCAGCGCGCTCTGTATATGCC 720
Db 661 GTCCCGCTCGCGGCTCCATCTCGGCGTCCGCGGACCTCAGCGCGCTCTGTATATGCC 720
QY GCGCGCATACCGGTCAACCGCGACGTCAAGGTTCAAGTTTGAAGAGGAAACCGAGAG 780
Db 721 GCGCGCATACCGGTCAACCGCGACGTCAAGGTTCAAGTTTGAAGAGGAAACCGAGAG 780
QY ATCATGTTTGGCGGAGGCGCATCTCGCTCTTGGTCTTCGAGGAGAGTCTTGGCGCG 840
Db 781 ATCATGTTTGGCGGAGGCGCATCTCGCTCTTGGTCTTCGAGGAGAGTCTTGGCGCG 840
QY AAGAGAGTCTCGTCTGATCAACGAAACGCGCGCTCCGCTCGATGAGACCTCAGT 900
Db 841 AAGAGAGTCTCGTCTGATCAACGAAACGCGCGCTCCGCTCGATGAGACCTCAGT 900
QY CTGCAAGATTCGCAATGCTCTCGCTCTCTCGAGGCGCTTGAAGGCTTCAAGGTGAG 960
Db 901 CTGCAAGATTCGCAATGCTCTCGCTCTCTCGAGGCGCTTGAAGGCTTCAAGGTGAG 960
QY GGCATGTCGCGCAGAGGCGCTCTCGCGCGCTTCAACGAGCTTCGCGCGCGCAGC 1020
Db 961 GGCATGTCGCGCAGAGGCGCTCTCGCGCGCTTCAACGAGCTTCGCGCGCGCAGC 1020
QY CCCGCGCAGGTGAGGTGCGCGGCAACATCCGCAAGCTCTTTCGCGCTGCTGCTGCGC 1080
Db 1021 CCCGCGCAGGTGAGGTGCGCGGCAACATCCGCAAGCTCTTTCGCGCTGCTGCTGCGC 1080
QY GTTGAAGCAGAGAGAGGTCAAGGTCAAGAGCAGAGGCGATTTCCGCGAGACCG 1140
Db 1081 GTTGAAGCAGAGAGAGGTCAAGGTCAAGAGCAGAGGCGATTTCCGCGAGACCG 1140
QY TACCGCTCCGACGCTGCTCAAGTTCTCGCGCGCTGCTGAGAGCATGATGACGCGC 1200
Db 1141 TACCGCTCCGACGCTGCTCAAGTTCTCGCGCGCTGCTGAGAGCATGATGACGCGC 1200
QY TACTGAGTCTTCTGCTCGAGAAACAACAACAACAACCGCGCTCTCGAGCTGAG 1260
Db 1201 TACTGAGTCTTCTGCTCGAGAAACAACAACAACAACCGCGCTCTCGAGCTGAG 1260
QY AACAGCAGACCGCGCAGCGCGCAACTTCCAGGCGTCCGCTGCTGATTTGATGAG 1320
Db 1261 AACAGCAGACCGCGCAGCGCGCAACTTCCAGGCGTCCGCTGCTGATTTGATGAG 1320
QY AAGACAGAGCTCGACCTCGCTCATCGGCAAGCTTCAAGTTCAAGAGTGCACCGAGTTG 1380
Db 1321 AAGACAGAGCTCGACCTCGCTCATCGGCAAGCTTCAAGTTCAAGAGTGCACCGAGTTG 1380
QY CTCAAGCTGACATGAACCGCGCGCTTCTGAGCTCGCTGCGAGACCGCTGCTC 1440
Db 1381 CTCAAGCTGACATGAACCGCGCGCTTCTGAGCTCGCTGCGAGACCGCTGCTC 1440

QY AACTATCACGGGAGGCGTTGAGCATTTGACATGCTGTGTTACGCTTGGAGCTCGGCAC 1500
Db 1441 AACTATCACGGGAGGCGTTGAGCATTTGACATGCTGTGTTACGCTTGGAGCTCGGCAC 1500
QY CTTGCGCAACCGCGTACTACCTTGTGCGAGCGCGCAGAGATGAGCAACAGGCGGTGAC 1560
Db 1501 CTTGCGCAACCGCGTACTACCTTGTGCGAGCGCGCAGAGATGAGCAACAGGCGGTGAC 1560
QY TCGCTCGCTCTCATTTCCGCGCGCGCGCATGCGGAGGCGCAACGAGCTCTTCTCTCT 1620
Db 1561 TCGCTCGCTCTCATTTCCGCGCGCGCGCATGCGGAGGCGCAACGAGCTCTTCTCTCT 1620
QY CTGCGCTCGACCTGTACTGACGCTCAGGCGGTGACCTCCGCGCGGAGTGAAGCTCGAC 1680
Db 1621 CTGCGCTCGACCTGTACTGACGCTCAGGCGGTGACCTCCGCGCGGAGTGAAGCTCGAC 1680
QY TTCAAGAGCAGTTGACCGCGCTTCTCCGAGCTCTCTCCAGAGACACTCGGCACTGAG 1740
Db 1681 TTCAAGAGCAGTTGACCGCGCTTCTCCGAGCTCTCTCCAGAGACACTCGGCACTGAG 1740
QY CTGAGCTGACAGCACTTGGCTGAGGTCAAGAGCGGCTCAACAGGCTCTTCGAGAG 1800
Db 1741 CTGAGCTGACAGCACTTGGCTGAGGTCAAGAGCGGCTCAACAGGCTCTTCGAGAG 1800
QY AAGAGCAGTACGACTTCGAGCGCGCTGAGCAGCAAGCTTCTGTAAGCGACCGGACAC 1860
Db 1801 AAGAGCAGTACGACTTCGAGCGCGCTGAGCAGCAAGCTTCTGTAAGCGACCGGACAC 1860
QY GTGCTGAGCTCTCTGCTGCTCTGCGCTCTGCGCAAGTCAACCTTACTGCGCTCAAG 1920
Db 1861 GTGCTGAGCTCTCTGCTGCTCTGCGCTCTGCGCAAGTCAACCTTACTGCGCTCAAG 1920
QY TGAAGGTTGCTCGCGCGGAGAGGCGCATCTGCTCAACGCGGAGGTGCGCAACCGCTTC 1980
Db 1921 TGAAGGTTGCTCGCGCGGAGAGGCGCATCTGCTCAACGCGGAGGTGCGCAACCGCTTC 1980
QY TGGAGAGCGCGCTTTCGAGGCGCGCGCGCAGCAATCTCTCGCGCGCAGCGCGCTC 2040
Db 1981 TGGAGAGCGCGCTTTCGAGGCGCGCGCGCAGCAATCTCTCGCGCGCAGCGCGCTC 2040
QY CTGTAAGTCTTCTGCGCGCGGAGAGCTCGGCTGAGAGCGCGCGCGCAGCTGTTGTC 2100
Db 2041 CTGTAAGTCTTCTGCGCGCGGAGAGCTCGGCTGAGAGCGCGCGCGCAGCTGTTGTC 2100
QY GCGGTGAGAGAGAGAGATCGGAGAGCAAGTCTCGCGCATCTACGAGGCGCATCAAGGAC 2160
Db 2101 GCGGTGAGAGAGAGAGATCGGAGAGCAAGTCTCGCGCATCTACGAGGCGCATCAAGGAC 2160
QY GCGCGCATCAACCAAGTCTCTGCTCAAGATGCTGCGCTTAAGGCCGAGCAAGCTCGCTA 2220
Db 2161 GCGCGCATCAACCAAGTCTCTGCTCAAGATGCTGCGCTTAAGGCCGAGCAAGCTCGCTA 2220
QY GAGCGCGCTCAACCCGAGACAGCTTTCGAGCGTCAAGGTGAGTGAAGAGTGAAGAGTCT 2280
Db 2221 GAGCGCGCTCAACCCGAGACAGCTTTCGAGCGTCAAGGTGAGTGAAGAGTGAAGAGTCT 2280
QY CTTCAATACAGATGCGCTTACTCTCTGCGCGTCAAGCTCTCAAGTTCTTCTGAT 2340
Db 2281 CTTCAATACAGATGCGCTTACTCTCTGCGCGTCAAGCTCTCAAGTTCTTCTGAT 2340
QY CCGCGCTCTCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
Db 2341 CCGCGCTCTCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
QY AGTTCAAAAAAAAAAAAA 2419
Db 2401 AGTTCAAAAAAAAAAAAA 2419

Patent No. 6355468
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
INVENTOR: Koopstra, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE OF INVENTION: Using Same
FILE REFERENCE: 29479/500NSC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 2475
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc difference
LOCATION: 113_34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
OTHER INFORMATION: Description of Artificial Sequence: Consensus
OTHER INFORMATION: Sequence of SEQ ID NOS: 12, 16, and 18
Patent No. 6355468
US-09-624-693A-20

Query Match 67.3%; Score 1629; DB 4; Length 2475;
Best Local Similarity 78.0%; Pred. No. 7,3e-307;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

QY 37 ATGCGCCCTTCCTTGAAGTCTGCTGCGCAGCAGCTGCGCAACGAGCTTTACCAACGCGCTCG 96
DB 1 ATGCGCCCTTCCTTGAAGTCTGCTGCGCAGCAGCTGCGCAACGAGCTTTACCAACGAGCTCG 60

QY 97 CAGCGCGCTCGCAGCAAGTCTGCG 154
DB 61 CAGCG 120

QY 155 GCCTGATGAGCGCAGCG 214
DB 121 GCTCGCTCTCTCCGACCGACCG 180

QY 215 ACCCGACCGAC---GACGTCGTGAGCTCAAGGAGGTACAGCTTCAAGCTTCAAGCTTCAAGCTT 271
DB 181 ACCCGACCGACCGACCG 240

QY 272 TCG 330
DB 241 TCG 300

QY 331 CGCGTGAAGAGAGCGTCACTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
DB 301 AAGATCGAAGAAAGTGAAGTCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360

QY 391 ACCAGCGGTTTGGTGGCTCG 450
DB 361 ACCAGCGGTTTGGTGGCTCG 420

QY 451 GCGCGTATGAGAGCGTCACTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510
DB 421 GCGCGTATGAGAGCGTCACTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480

QY 511 GAGCGCGCGCTCGAGAAACGCTTCTCGAGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
DB 481 GAGCGCGCGCTCGAGAAACGCTTCTCGAGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540

QY 571 GTCAACTCGCTCAAGCG 630
DB 541 GTCAACTCGCTCAAGCG 600

QY 631 AACTTCTGAACACCGCATCAAGCG 690
DB 601 AACTTCTGAACACCGCATCAAGCG 660

QY 691 GCGGACCTTCAAGCCCGCTCTGTAATCGCGCGCGCGCATGACCGGTCACCCCGACGTCAG 750
DB 661 GCGGACCTTCTCCCGCTCTGTAATCGCGCGCGCGCATGACCGGTCACCCCGACGTCAG 720

QY 751 GTTACGCTTTGACGAGGAGACCGAAGATCATTTTCCGCGCGAGCGCATCTTCTGCTC 810
DB 721 GTTACGCTTTGACGAGGAGACCGAAGATCATTTTCCGCGCGAGCGCATCTTCTGCTC 780

QY 811 TTTGCTCTGAGGAGATCT 870
DB 781 TTTGCTCTGAGGAGATCT 840

QY 871 GCGGCTCTCGGCTCTGATGAGGAGCGGCTCTGATGAGGAGCGGCTCTGATGAGGAGCGGCTCT 930
DB 841 GCGGCTCTCGGCTCTGATGAGGAGCGGCTCTGATGAGGAGCGGCTCTGATGAGGAGCGGCTCT 900

QY 931 TCGAGGCTTTGAGCGCTCTTACGCTGAGGAGCGATGTCGCGCGCGAGCGGCTCTTCTGCG 990
DB 901 TCGAGGCTTTGAGCGCTCTTACGCTGAGGAGCGATGTCGCGCGCGAGCGGCTCTTCTGCG 960

QY 991 CGGTTCAATCGACGAGCTCTGCG 1050
DB 961 CGGTTCAATCGACGAGCTCTGCG 1020

QY 1051 CGCAGCGCTCTTCTCGGCTCTGCTTCTGAGGAGCGAGGAGGTCGAAGTCTCAAG 1110
DB 1021 CGCAGCGCTCTTCTCGGCTCTGCTTCTGAGGAGCGAGGAGGTCGAAGTCTCAAG 1080

QY 1111 GACGACGAGGAGCTTTCTTCCGAGAGACCGGTAACCGGTCGCGAGCTGAGTCTGCTC 1170
DB 1081 GACGACGAGGAGCTTTCTTCCGAGAGACCGGTAACCGGTCGCGAGCTGAGTCTGCTC 1140

QY 1171 GCGCGCTCTGAGGAGATGATGACGCTTCTGATGAGGAGGTCGATGAGGAGGTCGATGAGGAG 1227
DB 1141 GCGCGCTCTGAGGAGATGATGACGCTTCTGATGAGGAGGTCGATGAGGAGGTCGATGAGGAG 1200

QY 1228 ACGAGACCGAACAACCGCTCTTCTGAGTGAAGAAAGAGACCGGCGAGCGGCGGAGC 1287
DB 1201 TCGAGACCGAACAACCGCTCTTCTGAGTGAAGAAAGAGACCGGCGAGCGGCGGAGC 1260

QY 1288 TTTCAAGGCGGCGCTGCTGATTTTCAATGAGAGAGACGAGGCTCGAGCTCGGCTCATC 1347
DB 1261 TTTCAAGGCGGCGCTGCTGATTTTCAATGAGAGAGACGAGGCTCGAGCTCGGCTCATC 1320

QY 1348 GCGAAGCTCAATTCGACGAGTGAACGAGTGTCTCAACGCTGCAATGAAACCGGCGCTG 1407
DB 1321 GCGAAGCTCAATTCGACGAGTGAACGAGTGTCTCAACGCTGCAATGAAACCGGCGCTG 1380

QY 1408 CTTTCTGCTCTGCTGCGAGAGACCGGCTCTCAATGATGAGGAGGCTTGAATG 1467
DB 1381 CTTTCTGCTCTGCTGCGAGAGACCGGCTCTCAATGATGAGGAGGCTTGAATG 1440

QY 1468 CACATGCTGCTTACGCTTGGAGTGGACCTTGGCAACCGGTCATCACTTCTGCTC 1527
DB 1441 GCGCGGCGGCTTACGCTTGGAGTGGACCTTGGCAACCGGTCATCACTTCTGCTC 1500

QY 1528 CAGCGCGAGAGTGGCAACCGGCTCTCAATGCTCTGCTCAATCTCGCGCGCGCGCGCGCG 1587
DB 1501 CAGCGCGAGAGTGGCAACCGGCTCTCAATGCTCTGCTCAATCTCGCGCGCGCGCGCGCGCG 1560

QY 1588 ACTGCGAGGAGCAAGAGTCTTCT 1647
DB 1561 ACTGCGAGGAGCAAGAGTCTTCT 1620

QY 1648 CAGGCGCTGACCTCTCGCGCGAGTGAAGTCACTTCAAGAAAGATTGACCGGCTTCTC 1707
DB 1621 CAGGCGCTGACCTCTCGCGCGAGTGAAGTCACTTCAAGAAAGATTGACCGGCTTCTC 1680

QY 1708 CCGACTCTCTCTCAAGAGACCTCTCGGACCTGCGCTCAAGCTC-----AAGCA 1755
DB 1681 CCGACTCTCTCTCAAGAGACCTCTCGGACCTGCGGCTCAAGCTC-----AAGCA 1740

QY 1756 CTTGCGCTGAGGTCGAAGAGCGCTCAAGAGCGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815

```

1  RESULT 3
2  US-09-627-216A-9
3  : Sequence 9, Application US/09627216A
4  : Patent No. 6368837
5  : GENERAL INFORMATION:
6  : APPLICANT: Saritslani, Sina F
7  : APPLICANT: Tang, Xiao-Song
8  : APPLICANT: Qi, Wei Wei
9  : APPLICANT: Vannelli, Todd
10 : APPLICANT: Gatenby, Anthony
11 : TITLE OF INVENTION: Biotransduction of para-Hydroxycinnamic Acid
12 : FILE REFERENCE: BC1009 US NA
13 : CURRENT APPLICATION NUMBER: US/09/627,216A
14 : CURRENT FILING DATE: 2000-07-27
15 : PRIOR APPLICATION NUMBER: 60/1147,719
16 : PRIOR FILING DATE: 1999-08-06
17 : NUMBER OF SEQ ID NOS: 14
18 : SOFTWARE: Microsoft Office 97
19 : SEQ ID NO 9
20 : LENGTH: 2151
21 : TYPE: DNA
22 : ORGANISM: Artificial Sequence
23 : FEATURE:
24 : OTHER INFORMATION: Description of Artificial Sequence: mutant from Rhodocortula glutinis

```

Query Match	49.5%	Score 1196.6	DB 4	Length 2151
Best Local Similarity	74.0%	Pred. No. 3.8e-223		
Matches 1565	Conservative 1	Mismatches 525	Indels 24	Gaps 3
QY	CGCGCTCCGACCAAGTCGGCTGGCGGCGCCACCTTCGGGCTCTCCGGCGACGCGCGGCGCT	158		
DB	45 CGTGCATCCGCAACAGCAAGCTGTCAATATGCGCCTTCACCAACCTTCGATCGCAGGCTC	104		
QY	159 CGATGGCCAGCGCCGCGACCAATCGACAGTCAGATCGTCAGAGCTCTCCAGCAGCC	218		
DB	105 GCACCTGCCCAACACCCAGGTACGACAGTGTGACATGTGCAAGAAATGTGCGCGCGC	164		
QY	219 CACCGACGAGTCGTGTGAGCTCAGCGGGTAAAGCTTACCGCTCCGTGAGTGTGTGGCGC	278		
DB	165 GACCTGACTCGACGCTCGAACTCGAGGCTATCCGCTCAACCTCGAGAGCGTGTCTCGGC	224		
QY	279 CGCGCCCAAGGGGCGACAGGTCCTCGGCTCAGAAAGACGAGATCCGCGCACGCGTCGA	338		
DB	225 CGGAGAGGAGGGGACAGGCTGTCCGCTTAAAGACAGAGATCCGCTCAAAGATTGA	284		
QY	339 CAAAGCGTCGACTTCTCTCAAGGCCAGCTTCAGAACTCGGTCTACGAGTCAACACGAG	398		
DB	285 CAATCGCTCGAGTCTTTCGCGCTCGCAACTCTCCATGAGCGCTTACGCGTCACAGCATGG	344		
QY	399 TTTCGGTGGCTGGCGCGACAGAGACATGAGAGATGACATGACGCTCCAAAGCGCTCAT	458		
DB	345 ATTTGGCGGATCCGACACCCGACCCGAGAGCGCATTTGCTTCAGAAAGCTCTTCT	404		
QY	459 CGAGCACCAAGCTCTCGGCGGTGAGCGCCAGAGTCGATCGTCTTCAGAGTGGACGCG	518		
DB	405 CGAGCACCAAGCTCTCGGCGGTGATTCCTTGTGTGTGACTGCTTCGCTCGCGCGCG	464		
QY	519 CCTCGAAGAACGCTTCCGCTCGAGGTGTTCGCGGCGCATAGTATCCGCTCACTC	578		
DB	465 TCTCGAAGACTCGCTTCCCTCGAGGTGTTCGCGCGCCATGACAAATCCGCTCAACG	524		
QY	579 GCTCAGCGGCGCACTGGCGCGCGGCTCGCTGTCTTGAAGGGCTTCACCAATCTT	638		
DB	525 CTTGACCGCGGCACTGCGCTGTGCGCTTCGTGCTTCGAGGGCTTCACCAATCTT	584		
QY	639 GAAACACCGCATACGCCCATGTCGCCCTCGCGGCTCCATCTCGCGTCCGCGGACCT	698		
DB	585 CAACACCGCATACGCCCATGTCGCCCTCGCGGCAATCTCGCGTCCGCGGACCT	644		
QY	699 CAGCCCGCTCTGTAAATTCGCGCGCGGCACTACCGGTACCCCGAGCTCAAGTTCAGT	758		
DB	645 CTCCTCTCTCTCAATATGAGGGGCAACAGGTCACCCGACACCAAGGTGACGT	704		
QY	759 TTTCGACAGAGGAACCGAAGATCAATGTTTGGCCGCGAGGCAATCTCGCTTTTGGCT	818		
DB	705 CGTTCACAGAGGGAAGAGAAATCTGTATGCGCGCGAGGGAGTGGCGCTTTCACACT	764		
QY	819 CAGGCAATGCTCTCGGCCGAAAGAGGTCTCGTCTGTCAACGAAAGCGCGCTCTC	878		
DB	765 CAGGCGCGTCTCTCGGCCGAAAGAGGTCTCGTCTGTCAACGCGACCGCGCTCTC	824		
QY	879 CGCTTCGATGCGACCTCAAGTCTGACAGCATGACATGCTGTGCTCTCTCGAGGCG	938		
DB	825 AGCATCATGCGACCTCGCTGTGACAGCAGCATGCTCTCTCTCGAGGCG	884		
QY	939 CTTGACGCTTCAAGTGAAGCATGTGCGGCGACAGGAGGCTGTTCGCGCGTTCAT	998		
DB	885 GCTTCAGCGCATGACGCTCGAACGATGTGCGGCAACGCGCGCTGTTCACCCCTTCT	944		
QY	999 CCAGACGCTTCCGCGCGCGACCGCGGCGAGGTGAGAGTTCGGCGCAATCCGACAGCT	1058		
DB	945 TCAAGAGCTCAAGCGGCTTCAACCGACGAGATCGAGTTCGGGGAAACATCCGCAAGCT	1004		
QY	1059 CCTTTCGGGCTGTGTTTCCGTTAGACAGAGAGAGGTCAAGTTCAGAGCGACGA	1118		
DB	1005 CCTCGAGGGAAGCGGCTTTGTGTCCACATATGAGAGAGGTTCAGAGTCAAGACGACGA	1064		

QY 1119 GGGCATTTCTTCCGAGAGACCGCTACCGCTCCGACGCTCGCTCACTTCTCCGCGCCGCT 1178
DB 1065 GGGCATTTCTTCCGAGAGACCGCTACCGCTCCGACGCTCGCTCACTTCTCCGCGCCGCT 1124
QY 1179 CGTGGAGACATGATGACGCTTACTGACTCTCTCGCTGGAACAACA---CGACGAC 1235
DB 1125 CGTCAGCAGCTCATTCAGAGCCGACGCTGCTTACCATGAGAGCCGCGCACTGACGAC 1184
QY 1236 CGACAGCCGCTCCCTCGAGCTGGAACAACAAGCAGCCGACGCGGCACTTCCAGGC 1295
DB 1185 CGACAGCCCTCTCATGACGCTGGAACAACAAGCAGCCGCGCACTTCCAGGC 1244
QY 1296 GTCCGCTGTGAGCAACACATGAGAGACTCGCTCGGCTCGCCGAGATCGCAAGCT 1304
DB 1305 CAATTCACGAGCTCAGCGAGTCTCAAGCGCGGCACTGAACCGCGGCTCCCTCTCTG 1364
QY 1416 CCTCGCTGCGAGAGACCGCTCGCTCACTATCAGCGCAAGGCTTGGACATTCATCGC 1475
DB 1365 CCTCGCGCGGAGAGACCGCTCGCTCTCTATCACTGCAAGGCTCGACATCGCGCTGC 1424
QY 1476 TCGTTACGTTTGGAGCTCGGCACTTGGCAACCGGCTCACTTCTCGTCCAGCGCGC 1535
DB 1425 GCGGTAACCTTGAGTTGGAGCACCTCGCACTCTGTGAGAGATGTCAGCCGCGC 1484
QY 1536 AGAGATGGGCAACCAAGCGCTCACTCGCTCGCTCACTCTCGCGCGCGCACTTGGCGA 1595
DB 1485 TGAAGATGGGCAACCAAGCGCTCACTCGCTCGCTCACTCTCGCGCGCGCACTTGGCGA 1544
QY 1596 GGGCAACGAGCTCTTCT 1555
DB 1545 GTCCAAACGAGCTCTTCT 1604
QY 1656 CGACCTCGCGGAGTGGAGCTGACCTCAAGAGAGTTGACCCGCTTCCCGAGCTCT 1715
DB 1605 CGACTTGGCGGAGCTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAG 1664
QY 1716 CCTCGAGCAGCCTCGGCACTGAGCTCGAAGCTCA-----ACGCACTTGGCT 1763
DB 1665 CATCGACGAGCCTTGGCTCGGCACTGAGCTCGAAGCTCGGCAAGCTCGTCA 1724
QY 1764 CGAGCTCAAGAGGCGCTCAAGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
DB 1725 GAAGGTGAACAAGAGCTGCGCAAGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1784
QY 1824 GCGCTGAGCAGAGCTTCTCTGTAAGCAGCAGCAGCTGAGCTCTCTCTCTCTCTCTCT 1883
DB 1785 GCGCTGAGCAGAGCTTCTCTGTAAGCAGCAGCAGCTGAGCTCTCTCTCTCTCTCTCT 1844
QY 1884 GCGCTGAGCAGAGCTTCTCTGTAAGCAGCAGCAGCTGAGCTCTCTCTCTCTCTCTCT 1943
DB 1845 GTC-----GCTCTGCTCGCGCGCTGCAAGCTGGAAGGCTGCGCGCGAGCTC 1895
QY 1944 GCGCATCTGAGTGAAGGAGAGGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2003
DB 1896 GCGCATCTGAGTGAAGGAGAGGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1955
QY 2004 GCGGAGCAGAGTACCTCTGCGAGCAGCAGCTGAGTCTGTAAGTCTGAGCGAGAG 2063
DB 1956 GCGGAGCAGAGTACCTCTGCGAGCAGCAGCTGAGTCTGTAAGTCTGAGCGAGAG 2015
QY 2064 GCTCGGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2123
DB 2016 GCTTGGGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2075
QY 2124 GAGAGAGTCTCGGAGTCAAGAGGAGTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2183
DB 2076 CTGGAAGCTTCCAAAGATCTAGAGAGGAGTCAAGTGGGAGAGATCAAGAGAGTCTCTCT 2135

QY 2184 CAAGATGCTGCGCTA 2198
DB 2136 CAAGATGCTGCGCTA 2150

RESULT 4
US-09-765-873A-9
Sequence 9, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BCI009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2151)
US-09-765-873A-9

Query Match 49.5%; Score 1196.6; DB 4; Length 2151;
Best Local Similarity 74.0%; Pred. No. 3.8e-223;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CCGCGCTCGACCAAGTGGGCTGGCGGCCCGCCACCTTGGGCTCTCCGCGAGCGCGGCT 158
DB 45 CGTGCATTCGGAAGAGAGGCTGCAATGCGCTTCAACCACTTCCGAGCTCGAGGCTC 104
QY 159 CGATGGCAGCGCGCGGAGCAGCTGAGCTGAGTGTGAGAGAGCTCTCAGCGAGCC 218
DB 105 GCACTGCGCAACACCGAGTCAAGAGAGTCAAGTGTGAGAGAGTGTGCGCGCC 164
QY 219 CACCAAGAGCTGCTGAGCTCAGGAGTCAAGCTCAGCTTCCGTAAGCTGTGCGGCG 278
DB 165 GACCACTGAGCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 224
QY 279 GCGCGCAAGGAGGAGCAGAGTCCGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
DB 225 CCGAG 284
QY 339 CAAAGAGCTGCACTTCTCAAGGAGCTTCAAGTCTGAGTCTGAGAGTCAACAGAG 398
DB 285 CAATGAGTCAAGTCTTGGCTGCAAGTCTTCAAGTCTGAGTCTGAGAGTCAACAGAG 344
QY 399 TTTGAGTGTGCTGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
DB 345 ATTTGAGAGTCAAG 404
QY 459 CGAGCAGAGCTCTGCGGAGTGAAGCGAGCTGAGCTGAGCTTCAAGCTTCAAGCTGAG 518
DB 405 CGAGCAGAGCTCTGCGGAGTGAAGCGAGCTGAGCTTCAAGCTTCAAGCTTCAAGCTGAG 464
QY 519 CTTGAGAGAGAGCTTCTGAGTGAAGCTGAGCTGAGCTTCAAGCTTCAAGCTTCAAGCTGAG 578
DB 465 TCTGAGAGAGAGCTTCTGAGTGAAGCTGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTGAG 524
QY 579 GCTCAAGAGTGAAGCTTCTGAGTGAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTGAG 638
DB 525 CTTGAGAGAGAGCTTCTGAGTGAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTGAG 584
QY 639 GAACAGAGAGTCAAG 698
DB 585 CAACAGAGAGTCAAG 644

QY 699 CAGCCGCTCTCGTACATCGCCGCGCCATCAACCGGTACCCCGACGTCAGATTACAGT 758
Db 645 CTCCTCTCTCTCTACATTGCAAGCGCCATCAAGCGGTACCCGGAAGAGTACAGT 704
QY 759 TTGGACAGAGGAAACCGAAGATCATGTTTGGCGCCGAGGCACTCTGCTTTTGGTCT 818
Db 705 GGTCCACAGAGGGAAGAGAGATCTGTACGCGCCGAGGAGATGGGCTCTTCAACCT 764
QY 819 CGAGGCAATGTCCTCGCCGCGAAGAGGATCTCGGTCTGATCAACGGAACGCGCGTCTC 878
Db 765 CGAGCCGCTCTGCTCGGCGCGAAGAGGATCTCGGTCTGATCAACGGAACGCGCGTCTC 824
QY 879 CGGCTGATGGCAACCTCTGATGTGCAACCTGCAATGCTCTGCTCTTCTGCAAGC 938
Db 825 AGCATGATGGCAACCTCTGATGTGCAACCTGCAATGCTCTGCTCTTCTGCAAGC 884
QY 939 CTGACGAGCTCTGACGATGGAAGGCAATGGTGGGCGAAGAGGCTGTTGCGCGCTTCA 998
Db 885 GCTCAGGCGCATACGCTGGAAGCATGATGGGCAACGCGCTGTTCTCAACCTTCTCT 944
QY 999 CACGACGTCGTGCGCCGCGACCCCGGAGGTGAGGTGCGCGCAACATCCGACGCT 1058
Db 945 TCACGACGTCACGCGCCCTCACCCGACGATGCAAGTGGGGAACATCCGCAAGCT 1004
QY 1059 CCTTCCGCTGCTGCTGTTGGGTTGAGACGAGAGAGATCAAGGTCAGAGACGACGA 1118
Db 1005 CTTGAGAGGAAGCCGCTTGTGTCTCACCATGAGAGAGATCAAGGTCAGAGACGACGA 1064
QY 1119 GGGCATTTCTTGGCAGAGACCGCTAACCGCTCCGACGCTGACGCTCACTTCTGACGCGCT 1178
Db 1065 GGGCATTTCTTGGCAGAGACCGCTAACCGCTCCGACGCTGACGCTCACTTCTGACGCGCT 1124
QY 1179 CGTGGAGACATGATGACAGCTACTTCACTTCTGCTGCAAGACAA---CGACGAC 1235
Db 1125 CGTACGACGCTCATTTACGCGCCACGCGCTCTCTCACTACGAGGCGCGGCACTGACGAC 1184
QY 1236 CGACACCCGCTCTCTGACGTCGAGAACAGACAGACGCGCGGCAACTTCCAGGCG 1295
Db 1185 CGACACCCGCTCTCTGACGTCGAGAACAGACGCTTCCAGACGCGCGGCAACTTCCAGGCG 1244
QY 1296 GTGCGGTGTCGATTTGATGAGAGAACAGGCTGACGCTCCCTCACTGCGACGCT 1355
Db 1245 TGCCTGTGTGGCCAAACCATGAGAGAACTGCTGGGCTGCCCAAGTCCGCAAGCT 1304
QY 1356 CAACCTCAACGAGTGAACGAGTTGCTCAAGCTGCAATGAACCGCGGCTGCTTGTG 1415
Db 1305 CAACCTCAACGAGTGAACGAGTTGCTCAAGCTGCAATGAACCGCGGCTGCTTGTG 1364
QY 1416 CTTGCTGCGAGAGACCCGCTGCTCACTATCAACGAGAGGCTTGAATTCACATCGC 1475
Db 1365 CTTGCTGCGAGAGACCCGCTGCTCACTATCAACGAGAGGCTTGAATTCACATCGC 1424
QY 1476 TGTCTTACGTTTGGAGCTGCGCACTTGGCAACCCGCTCACTTCTGCAAGCCG 1535
Db 1425 GGTCTACACTCTGGAGTTGGAGACACTTGGCAACCCGCTGCAAGCTTCTGCAAGCCG 1484
QY 1536 AGAGATGGGCAACGAGGCGCTCAACTCGCTGCTATCTCCGCGCGCGCACTGCGCA 1595
Db 1485 TGAGATGGGCAACGAGGCGCTCAACTCGCTGCTATCTCCGCGCGCGCACTGCGCA 1544
QY 1596 GGCAGACGATCTTTTCTTCTCTTCTGCGCTGCACTGATGCAAGGCTGCAAGGCGCT 1655
Db 1545 GTCCACGACGATCTTTTCTTCTCTTCTGCGCAACCTGATGCTTCTCAAGACCAT 1604
QY 1656 CGACCTCGCGGATGAGCTGCACTTCAAGAAACATGAGCCGCTTCTCCGACGCT 1715
Db 1605 CGACCTCGCGGATGAGCTGCACTTCAAGAAACATGAGCCGCTTCTCCGACGCT 1664
QY 1716 CTTCCAGACGACTCTGCGACGCTGCAAGTCA-----ACGCACTTGGCT 1763
Db 1665 CATGACACGACATTTGGCTCGCGCATGACGCGGCTGCAACCTGCGGACGAGCTGCTGA 1724

QY 1764 CGAGTCAAGAGGCGCTCAACAGGCTCTGACAGACGACGATACGACTGAGCC 1823
Db 1725 GAAGGTGAACAGAGCTGCGCAAGGCTCTGAGAGACCACTGTAAGACTGCTGCC 1784
QY 1824 GCGCTGGCAAGACGCTTCTGTGATGCGACCGGCAACGCTGCTGCTTCTGCTCT 1883
Db 1785 GCGCTGGCAAGACGCTTCTGTGATGCGACCGGCAACGCTGCTGCTTCTGCTCT 1844
QY 1884 GCGCTTGGCAACGCTTCTGTGATGCGACCGGCAACGCTTCTGCTTCTGCTCT 1943
Db 1845 GTC-----GCTTCTGCTGCGCGCTGCAAGGCTTCTGCTGCTGCTGCT 1895
QY 1944 GGCATTTCTGCTCAACGCGGAGGTGCGCAACGCTTCTGCAAGACGCTTCTGCAAGC 2003
Db 1896 GGCATTTCTGCTCAACGCGGAGGTGCGCAACGCTTCTGCAAGACGCTTCTGCAAGC 1955
QY 2004 GCGCGCGACGATCTCTGCGCGGACGCGGCTCTGATCTGCTGCTGCGGACGA 2063
Db 1956 GCGCGCGCTCTGATCTCTGCGCGGACGCGGCTCTGATCTGCTGCTGCGGACGA 2015
QY 2064 GCTCGCGGTGAGAGGCGCGCGGCGACGATGTTGTGCGGCTGAGAGACGATCGG 2123
Db 2016 GCTTGGCTTCAAGGCGCGCGGCGGAGAGCTTCTCTGCAAGACGATGAGATCGG 2075
QY 2124 GAGCAAGCTCTGCGGATCTTGAAGGCTCAAGAGACGCGGATCAACGCTCTCT 2183
Db 2076 CTGCAACGCTTCAAGATCTTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCT 2135
QY 2184 CAAGATCTGCGCTA 2198
Db 2136 CAAGATCTGCGCTA 2150

RESULT 5
US-09-624-693A-18
; Sequence 18, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocistara, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/50NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Rhodotorula toruloides
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ..(2148)
US-09-624-693A-18

Query Match 49.5%; Score 1196.6; DB 4; Length 2439;
Best Local Similarity 74.0%; Pred. No. 3; 9e-223;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCCAGCAAGTGGCTGCGGCGCCCACTTGGCTCTCCGCGGACGCGCGCT 158
Db 45 CGTGGATTCGCAAGAGGCTGTCAATGAGCGCTGCAACCTTGGAGTGGAGGCTC 104
QY 159 CGATGGCAAGCGCGCGGCAAGTGGCTGCAATGATGTGCAAGGCTCTTGAAGGACCC 218
Db 105 GCACTTGGCCCAACCAAGGTGCAAGGCTGCAATGATGTGCAAGGCTCTTGAAGGACCC 164
QY 219 CACGACGAGCTGATGAGCTGAGCGGTACAGCTCAACGCTGAGCTGAGTGTGCGGCG 278
Db 165 GACGACGAGCTGAGCTGAGCTGAGCGGTACAGCTCAACGCTGAGCTGAGTGTGCGGCG 224

279 CGCCGCAAGGGGCGAGGGTCCGCTCCAGAAAGAGAGATCCGCGAGCTGCA 338
225 CGCAGAGAGGGGCGAGCTGTCCGCTCAAGAGAGAGAGATCCGCTCAAAAGTTGA 284
339 CAAGAGCGTCACTTCTCTCAAGGCCAGCTTCAGAACTCGGTTCAGAGTCAACAAGG 398
285 CAATTCGGTCAAGTTCTTGCGCTGCAACTCTCATAGCGGTCTACGGGTCAAGCTGG 344
399 TTTCGGTGGCTCGGCGCAGACAGAGAGATGCAAGTCAAGCTCCAGAGGCGCTCAT 458
345 ATTGGCGAGATCCGAGACACCCGACCGAGAGCCATCTCGCTTCAGAAAGCTTCTCT 404
459 CGAGCAGCAGCTTCGCGAGCTGAGCGGCGAGCTGCTGCTTCAGAGCTGAGCGAG 518
405 CGAGGACAGCTTCGCGAGTTCCTCCCTCGTTCGATCTCGCTCGCGCGCGG 464
519 CCTGAGAAACGCTTCGCTGAGAGTGTCCGCGGCGCATGTATCCGCTCAACTC 578
465 TCTCAGAACTCGCTTCCCTCGAGGTTGTTCGGCGCATGACAACTCGCTCAACAG 524
579 GCTCAGCGTGGCCACTCGCGCGCTGCTGCTTCGATGAGGCGCTCAACAACTCTT 638
525 CTGACCCGCGGCGCACTGCGCTGTCCGCTGTGCTCTGAGGCGCTCAACAACTTCT 584
639 GAACCAACGATCAAGCCCATGCTCCCTTCGCGCTCATCTCGCGCTGCGCGAGCT 698
585 CAACCAACGATCAACCCCATGCTCCCTTCGCGCGCATCTCTGCGCGCGAGCT 644
699 CAGCCCGCTTCGATCATGCGCGGCGCATCAGCGGTCAACCCGAGCTCAAGGTTCACT 758
645 CTCTCTCTCTCTCAATTGAGCGGCGCATAGGGGTACCCGAGCAAGAGTGAAGCT 704
759 TTTCAGCAGAGGAGCCGAGAGATCATTTTTCGCGCGAGGCGCATCTGCTCTTTGCT 818
705 CGTCAACGAGGCGAGAGAGATCTGTACGCCGCGAGGCGATGCGCTCTTCAACCT 764
819 CGAGGAGTGTCTCTCGCGCGCGAGAGAGAGTCTGCTGTCAACGAGAGCGCGCTCT 878
765 CGAGCGCGTGTCTCTCGCGCGAGAGAGTCTGCTGTCAACGAGAGCGCGCTCT 824
879 CGCCTCGATGAGGAGCCCTCACTGTCAGAGCTGCGACATGCTCTGCTCTCGAGG 938
825 AGCATGATGCGCACTCTGCTGTGCAAGAGCAACATGCTCTGCTCTCTGAGGTC 884
939 CTTCAGCGCTCTCAGCGTGAAGGCGCATGTGCGCGAGAGGCTCTGTCGCGCTCAT 998
885 GCTCAAGCGCATGAGCGTCAAGCGATGATGCGCGACCGCGCTCTTCAACCTTCTCT 944
999 CACGACGCTTCGCGCGCGAGACCCGCGCGAGTGAAGTTCGCGCGCAACATCCGACGCT 1058
945 TACGAGCGTCAAGCGCGCTCAACCGAGAGATGAGATTCGCGGAGAGATCCGCAAGCT 1004
1059 CTTTTCCGCTCGTCTGTTTCCGTTGAGCAAGAGAGAGAGTCAAGGTCAGAGAGAG 1118
1005 CTTGAGAGGAGAGCGCTTGTGTCAACATGAGAGAGAGTCAAGGTCAGAGAGAG 1064
1119 GGGCATTTCTTCGAG 1178
1065 GGGCATTTCTTCGAG 1124
1179 CGTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1235
1125 CGTCAAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
1185 CGAG 1244
1296 GTGCGCTGTCTGATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
1245 TCGCGCTGTGAG 1304
1356 CAACCTTACGAG 1415

1364
1305 CAACCTTACGAG 1364
1416 CTTGCGCGAG 1475
1365 CTTGCGCGAG 1424
1476 TGTTCAGCTTCGAG 1535
1425 GAGCTTACCTTCGAG 1484
1536 AGAGATGAG 1595
1485 TGAATGAG 1544
1596 GAGCAG 1655
1545 GTCAAG 1604
1656 GAGCTTCGCGAG 1715
1605 GAGCTTCGCGAG 1664
1716 CTTGAG 1763
1665 CATGAG 1724
1764 CAGAGTCAAG 1823
1725 GAAGTGAAG 1784
1824 GCGTGGAG 1883
1785 GCGTGGAG 1844
1884 GCGTGGAG 1943
1845 GTC-----GCTTCGCTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1895
1944 GCGCATCTTCGCTCAAG 2003
1896 GCGCATCTTCGCTCAAG 1955
2004 GCGGCGAG 2063
1956 GCGGCGAG 2015
2064 GCTGCGGTGAG 2123
2016 GCTGCGGTGAG 2075
2124 GAGCAG 2183
2076 CTGAG 2135
2184 CAAGATGCTCGGCTA 2198
2136 CAAGATGCTCGGCTA 2150

RESULT 6
US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 636837
; GENERAL INFORMATION:
; APPLICANT: Sarsisani, Sina F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei-Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bio-production of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A

CURRENT FILING DATE: 2000-07-27
 PRIOR APPLICATION NUMBER: 60/147,719
 PRIOR FILING DATE: 1999-08-06
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: Microsoft Office 97
 SEQ ID NO: 7
 LENGTH: 2151
 TYPE: DNA
 ORGANISM: Rhodococcus glutinis
 US-09-627-216A-7

Query Match 49.4%; Score 1195; DB 4; Length 2151;
 Best Local Similarity 73.9%; Pred. No. 7, 8e-223;
 Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

QY 99 CGCCGCTCCGACCAAGTGGCTGGGGCCCACTTGGGCTTCGCGCCGACGCCGCT 158
 DB 45 CGTGGATCGCAAAACAGGCTGTCAATGGCCCTCGACCAACTCGCATCGAGGCTC 104
 QY 159 CGATGGCCACGCGCGCACCAAGTCGAGCTCGAGATTCGAGAGAGAGGCTCTCAGAGACC 218
 DB 105 GCACCTGCCCAACACCCAGGTACAGGATCGAATGTCGAGAAAGATGCTGCGCGCC 164
 QY 219 CACCGACAGCTGTGAGCTCAGCGGATACAGCTCAACGTCGCTGAGCTTGTGCGCG 278
 DB 165 GACCGACTCGACGCTGAACTCGACGCTACTGCTCAACCTCGGAGACGTGCTCGGC 224
 QY 279 CGCCCGCAAGGGGCGGAGGCTCGGCTCCAGAACGACGACGAGATCCGCGACCGCTCA 338
 DB 225 CGGAGAAAGGAGGAGGCTGTCCGGTCAAGGACGACGACGATCGCTCAAGATTGA 284
 QY 339 CAAAGACGTGCACTTCTCAAGGCCAGCTTCAGAACTCGGTCTACGAGATCAACAGG 398
 DB 285 CAATCGGTGCAATTCCTTCCGCTCGCACTCTCCATGAGGCTTACGGCGTCAAGCTCG 344
 QY 399 TTTCGGTGGCTGGCGGACAGAGATGAGATGCAATCAAGCTCCAGAAAGCGCTCAT 458
 DB 345 ATTGGCGGATCCGAGACACCGGACGAGAGACCCCATTCGCTCCAGAAAGGCTCTCT 404
 QY 459 CGAGCACAGCTCTCGGCGGTGACGCGACGTCGCTCGCTTCAGCGCTCGGAGCGG 518
 DB 405 CGAGCACAGCTCTCGGCGGTGACGCGACGTCGCTCGCTTCAGCGCTCGGAGCGG 464
 QY 519 CCTCGAAGACAGCTTCCGCTGAGTGTGTCCGCGGCGCATGCTATCCGCTCAATC 578
 DB 465 TCTCGAAGACAGCTTCCGCTGAGTGTGTCCGCGGCGCATGCTATCCGCTCAATC 524
 QY 579 GCTCAAGCGTGGGCACTCGGCGGTGACGCGACGTCGCTCGCTTCAGCGCTCGGAGCGG 638
 DB 525 CTTGACCGCGGCGCACTCGGCGGTGACGCGACGTCGCTCGCTTCAGCGCTCGGAGCGG 584
 QY 639 GAAACACCGCATCAGCGCCATGTCCTCCCTCCGCGGCTCCATCTCGCGCTCGGAGCGG 698
 DB 585 CAACACCGCATCAGCGCCATGTCCTCCCTCCGCGGCGCATCTCGCGCTCGGAGCGG 644
 QY 699 CAGCGCGCTCTCGTCAATCGCGCGGCGCATCAGCGGTCAACCGGACGCTCAAGCTT 758
 DB 645 GTCTCTCTCTCTCAATTCGAGCGCGCATCAGCGGTCAACCGGACGCTCAAGCTT 704
 QY 759 TTTCGACGAGGAAACGAGAGATCATGTTGCGCGGAGGAGGAGGAGGAGGAGGAGG 818
 DB 705 CGTCAACGAGGAAACGAGAGATCATGTTGCGCGGAGGAGGAGGAGGAGGAGGAGG 764
 QY 819 CGAGGCAATGCTCTCTCGGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 878
 DB 765 CGAGGCGCTGTCTCTCGGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 824
 QY 879 CGCCTGATGCGGACCTCTGAGTGTGACGAGCTCGACATGCTCTGCTCTCTCGAGG 938
 DB 825 AGCATCATGCGGACCTCTGAGTGTGACGAGCTCGACATGCTCTGCTCTCTCGAGG 884
 QY 939 CTTAGCGCTCTCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 998

DB 885 GCTCAGGCGCATGAGCGGTGAAAGCGATGTCGGGCCACCGGCTGTTCCACCCCTTCT 944
 QY 999 CCAGCAGCTGTGCGCGCGGACACCGCGGACAGGTGAGGTCGCGGCAACATCCGACGCT 1058
 DB 945 TACGACGCTCAGCGCGCTTCAACCGGACGAGCATGAAAGTGGGGGAAACATCCGACGCT 1004
 QY 1059 CTTTCCGCTGTGCTTGTGCTTGAACAGAGAGAGGATCAAGGTCAAGAGACGAG 1118
 DB 1005 CTTGAGAGGAAAGCCCTTGTGCTTGAACAGAGAGAGGATCAAGGTCAAGAGACGAG 1064
 QY 1119 GGGCAATCTTGTGCGGAGACCGCTACCGGCTCCGACAGTGGCTTCAATTCCTCGCGCT 1178
 DB 1065 GAGCATCTCGCGGAGACCGCTACCGGCTTGGGACAGCTTCTCAAGTGGCTCGCGCT 1124
 QY 1179 CGTGGAGGACATGATGACAGGCTTACGACCTCTGCTCGAGAACAAACA---GACGAC 1235
 DB 1125 CGTCAAGCAGCTCAATTCAGCGCGCAAGCGGCTTCAACATGAGGCGCGGACAGTCAAG 1184
 QY 1236 CGACAAACCGCTCTCTGACGCTGAGAACAGACAGACCGGCAAGCGGCAATTCCAGGCT 1295
 DB 1185 CGACAAACCGCTCTCTGACGCTGAGAACAGACAGACCGGCAAGCTTCCAGGCT 1244
 QY 1296 GTGCGCTGTCTGATTTGATGAGAAACAGGCTGCACTGCGCTCATTCGCAAGCT 1355
 DB 1245 TGCGCTGTGCGCAACACAGAGAGAGAGCTCGCTTGGGCTGCGGCAATTCGCAAGCT 1304
 QY 1356 CAATTCACGAGTGCACCGAGTGTCTCAACGCTGACATGACCGCGCTGCTTCTGT 1415
 DB 1305 CAATTCACGAGTGCACCGAGTGTCTCAACGCTGACATGACCGCGCTGCTTCTGT 1364
 QY 1416 CTTGCTGCGGAGACCGGCTGCTCACTTCAACGAGAGGAGGAGGAGGAGGAGGAGGAG 1475
 DB 1365 CTTGCTGCGGAGACCGGCTGCTCACTTCAACGAGAGGAGGAGGAGGAGGAGGAGGAG 1424
 QY 1476 TGTTCAGCTTGGAGTGGGCACTTGTGCAACCGGCTCACTTCACTTGTTCAGGCGCG 1535
 DB 1425 GCGGTACACCTCGGAGTGGGCACTTGTGCAACCGGCTCACTTGTTCAGGCGCG 1484
 QY 1536 AGAGATGGGACACGAGGCGGTCACTTGTGCTTCACTTGTTCAGGCGCG 1595
 DB 1485 TGAATGGGCAACGAGGCGGTCACTTGTGCTTCACTTGTTCAGGCGCG 1544
 QY 1596 GCGCAAGAGCTCTTCT 1655
 DB 1545 GTCCACAGAGTCTTCT 1604
 QY 1656 GAGCTTGGCGGAGTGAAGTCTTCAAGAAAGAGTTCAGACCGGCTTCTCTCTCTCTCTCT 1715
 DB 1605 GAGCTTGGCGGAGTGAAGTCTTCAAGAAAGAGTTCAGACCGGCTTCTCTCTCTCTCTCT 1664
 QY 1716 CTTGAGAGAGCTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1763
 DB 1665 CATGAGACAGAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1724
 QY 1764 CAGGCTCAAGAGCGCTTCAACAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
 DB 1725 GAGGCTCAAGAGCGCTTCAACAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1784
 QY 1824 GCGGTGAGAGAGCGCTTCAAG 1883
 DB 1785 GCGGTGAGAGAGCGCTTCAAG 1844
 QY 1884 GCGCTTGGCCAGAGTCAAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
 DB 1845 GTC-----GCTCTGCTGCGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1895
 QY 1944 GGCATCTGCTTCAAG 2003
 DB 1896 GGCATCTGCTTCAAG 1955
 QY 2004 GCGGCGGACGATACCTCTGCGCGGACAGGCGGCTCTGATCTGATCTGATCTGATCTGATCT 2063
 DB 1956 GCGGCGGCTCTGATCTCTGCGCGGACAGTCAAGTCTTCAAGAGAGAGAGAGAGAGAGAG 2015

QY 1656 CGACCTCCGCGAGTGAAGTGCATCTTCAAGAGAGAGTTGACCCGCTTCTCCGACTCT 1715
DB 1605 CGACTTCGCGGCATGAGTTGAGTTCAAGAAAGCAAGTTCCGAGCCATCTGTCGCT 1664
QY 1716 CCTCCAGCAGACCTCGGCACTGGCTTCGACGTCA-----ACGCATTGGCGCT 1763
DB 1665 CATCAGCAGACCTTTGGCTCGGCATGACCGGCTCGAACTCGCGCAGAGCTTCGCA 1724
QY 1764 CGAGTTCAGAAAGCGCTCAACAAGGTTTCAGAGAGAGAGAGTACGACTTCGAGCC 1823
DB 1725 GAAGGTGAACAAACCTCGCCCAAGCGCTCGAGAGACCAACTCTGTCGACTCTGCC 1784
QY 1824 GCGCTGAGCAGACGCTTCTGTAAGCAGCGGACCGTCTGAGCTCTCTGCTCTC 1883
DB 1785 GCGCTGAGCAGAGCTTCTCTCTTCCGCGCGGACCGTCTGAGAGTCTCTGTCGAC 1844
QY 1884 GCGCTTGCACAGTACCTTACCTTACCTGTCACGCGTGAAGGTTGCTCGCCGAGAA 1943
DB 1845 GTC-----GCTCTCGCTCGCGCGCTCAACGCTGAAGGTGCGCGCGCGAGTC 1895
QY 1944 GGCATCTCGCTCAAGCGGAGAGTGGGAACCGCTTCGAGAGAGCGCTTTCGAGGC 2003
DB 1896 GGCATCTCGCTCAAGCGGAGAGTGGGAACCGCTTCGAGAGAGCGCTTTCGAGGC 1955
QY 2004 GCGGAGCAGAGTACCTTCGCGCGGACGCGCTCTGACTGTTCTGCGCGAGGA 2063
DB 1956 GCGCGGCTCTGTAAGCTTCGCGCGGACGCTCAAGTCTTCTGCTTCGCGAGGA 2015
QY 2064 GCTCGCGCTGAGAGGCGCGCGGAGAGTGTGTTGCTGCGCTGAGAGAGAGTCCG 2123
DB 2016 GCTGCGCTGAGAGGCGCGCGGAGAGTGTGTTGCTGCGCTGAGAGAGAGTCCG 2075
QY 2124 GAGCAAGCTCTCGGCATCTACAGAGGAGCATCAAGAGAGCGGATCAACAGCTCCG 2183
DB 2076 CTCGAAGCTCTCAAGATCTAGAGGAGCATCAAGTCCGCGGAGATCAACAGCTCC 2135
QY 2184 CAAGATGCTCGCTA 2198
DB 2136 CAAGATGCTCGCTA 2150

RESULT 8

US-09-765-873A-31
Sequence 31, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BCI009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 31
LENGTH: 2061
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS1
LOCATION: (1)..(2061)
US-09-765-873A-31

Query Match 49.1%; Score 1188.8; DB 4; Length 2061;
Best Local Similarity 75.2%; Pred. No. 1.2e-221;
Matches 1529; Conservative 1; Mismatches 478; Indels 24; Gaps 3;
QY 182 CGCAGCTCGAGATGTGCAGAGAGTCTCTCAGCAGCCCAACGAGAGCTGTGAGCTCA 241
|||||
DB 118 AGAACAAGACTTCGACCAAGCGGCAATTTTCAGAGCTGCGGCTGAGCAACATG 1177

DB 38 CGCAGTGCATCTCTGAGAAATGCTCGCGCGGACCGACTGACCTGCACTCG 97
QY 242 GCGGATACAGCTTCAAGCTTCGCTGAGAGTGTGCGGCGCGCCGCAAGAGGAGCGAGGTC 301
DB 98 ACGGCTACCTCGCTCAAGCTTCGAGAGAGTGTGCTGCGCGGAGAGAGGAGGCTGTCC 157
QY 302 GCGTTCAGAGAGAGAGAGAGATTCGCGGAGCGCTGCAAGAGAGCTGACTTCTTAAG 361
DB 158 GCGTTCAGAGAGAGAGAGATTCGCTCAAGATTAAGTCAATCGTTCGAGTCTTCGCT 217
QY 362 CCGAGCTTCAGAACTCGGCTTACAGAGTCAAGAGGTTTCGAGGCTCGGCGAGAGCA 421
DB 218 CGCACTCTCATAGAGCTTCAAGCGCTTCAAGAGCTTCAAGAGCTTCAAGAGCTTCA 277
QY 422 GAGTTCAGAGATGAGTCAAGCTTCAAGAGGCGCTTCAAGAGAGAGCTTCGAGGCTGA 481
DB 278 GCACCGAGAGAGCGCATCTCGCTCCAGAAAGCTCTCCAGAGAGAGCTTCGAGGCTG 337
QY 482 GCGCAGAGTCCGCTTCGCTTCAAGGTCGAGAGGCGGCTTCGAGAGAGAGCTTCGCTG 541
DB 338 TCCCTTCGCTTCAAGCTTTCGCTTCGAGAGGCTTCGAGAGAGAGCTTCGCTGCTG 397
QY 342 AGGTTCGCGGAGCGCATGATTCAGCGCTCAAGCTTCAGAGAGAGAGCTTCGAGGCTG 601
DB 398 AGGTTCGCGGAGCGCATGATTCAGCGCTCAAGCTTCAGAGAGAGAGCTTCGAGGCTG 457
QY 602 TCCGCTTCGCTTCGCTTCAAGAGGCTTCAGAGAGAGAGCTTCGAGAGAGAGCTTCG 661
DB 458 TCCGCTTCGCTTCGCTTCAAGAGGCTTCAGAGAGAGAGCTTCGAGAGAGAGCTTCG 517
QY 662 TCCGCTTCGCGGCTTCATCTCGGCTTCGAGAGAGAGCTTCAGAGAGAGAGCTTCG 721
DB 518 TCCGCTTCGCGGCTTCATCTCGGCTTCGAGAGAGAGCTTCAGAGAGAGAGCTTCG 577
QY 722 GCGCATACAGCTTCAGAGAGAGAGCTTCAGAGAGAGAGCTTCAGAGAGAGAGCTTCG 781
DB 578 GCGCATACAGCTTCAGAGAGAGAGCTTCAGAGAGAGAGCTTCAGAGAGAGAGCTTCG 637
QY 782 TCATGTTGCGGAGAGAGAGAGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGCTTC 841
DB 638 TCCGTACAGCGCGGAGAGAGAGAGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAG 697
QY 842 AGAGAGGCTTCGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 901
DB 698 AGAGAGGCTTCGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 757
QY 902 TGCAAGAGCTTCGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 961
DB 758 TGCAAGAGCTTCGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 817
QY 962 CCAATGCTCGCAGAGAGAGAGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGCTTC 1021
DB 818 CCAATGCTCGCAGAGAGAGAGCTTCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGCTTC 877
QY 1022 CCGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGCTTC 1081
DB 878 CCGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGCTTCG 937
QY 1082 TTGAGCAGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1141
DB 938 TTGAGCAGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 997
QY 1142 ACCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1201
DB 998 ACCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1057
QY 1202 ACTGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1258
DB 1058 ACCGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1117
QY 1259 AGAACAAGAGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAG 1318
DB 1118 AGAACAAGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAGAGAGCTTCGAGAGAG 1177

2Y 1319 AGAAGACAGGCTTGCACTGGCCCTCATGGGAGAGTCAATTGACGACGACCGAGT 1378
Db 1178 AGAAGACTGGCTTGAGGCTGGCCCAATGGCAAGCTCACTTACGACGCTACGAGAA 1237
2Y 1379 TGCTCAAGCTGACCATGAAACGCGGCTGCTTGTGCTCGCTGCGAGAGACCCGCTGC 1438
Db 1238 TGCTCAAGCTGACCATGAAACGCGGCTGCTTGTGCTCGCTGCGAGAGACCCGCTGC 1297
2Y 1439 TCAACTATACGAGGAGGCTTGAACATTCATGATCGCTGCTTACGCTTGAGAGCTGCGC 1498
Db 1298 TCTCTACTCACTGCAAGGCTCTGACATGCGCTGCGGAGTACCTGAGAGTGGAGC 1357
2Y 1499 ACCCTGGCAACCGGCTGACATCTTGTGACGCGCGGAGAGATGGGAAACGAGCGCTCA 1558
Db 1358 ACCTGGCAACCGCTGAGACGATGCTTCAAGCGGCTGAGATGGGAAACGAGCGCTCA 1417
2Y 1559 ACTGCTGCTCTGATCTCGGCGCGGCGCACTGCGAGGCGCAACGAGCTCTTCTCTCC 1618
Db 1418 ACTGCTGCTCTGATCTCGGCTGCTGCGAGACGAGCTCAACGAGCTCTTCTCTCC 1477
2Y 1619 TTCTGCTGCTGACCTGATCTGACGAGCTGCAAGGCGGCTGCACTGCGGCGGAGAGAGCTCG 1678
Db 1478 TTCTGCTGCTGACCTGATCTGAGCTGCTTCAAGGCTGACCTTGGCGGAGTGAAGTTCG 1537
2Y 1679 ACTTCAAGACAGTTCGACCCGCTTCTCCGACTCTCTCTCAAGACGACCTTGGGACTG 1738
Db 1538 AGTTCAAGACAGTTCGAGCCGAGCATCTGCTGCTCATGACGACGACTTGGCTCG 1597
2Y 1739 GCTTCGACG-----TCAAGCACTTGGCTGAGGTGAAGAGGCTTCAAC 1786
Db 1598 CCATGACCGGCTGCAACCTGCGCGAGGCTGCTGCGAGAGTGAACAAACGCTGCGCA 1657
2Y 1787 AGGCTGCGAGAGAGAGAGAGAGTACGAGCTGAGAGCGCGGCTGAGAGCGCTTCTCTG 1846
Db 1658 AGGCTGCGAGAGAGAGAGAGTACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1717
2Y 1847 AGGACGCGGAGAGAGAGAGTACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
Db 1718 TCGCGCGCGGAGAGAGTACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1768
2Y 1907 CTGCGCTCAAGCGCTGAGAGAGTTCGCTGCGGAGAGGCGCATCTGCTACGCGGAG 1966
Db 1769 CGCGCTGCAAGCGCTGAGAGAGTTCGCTGCGGAGAGGCGCATCTGCTACGCGGAG 1828
2Y 1967 TCGGCAAGCGCTTCTGAGAGAGCGCGCTTCTGCGAGAGCGCGCGGAGAGTACCTCTCG 2026
Db 1828 TCGGCAAGCGCTTCTGAGAGAGCGCGCTTCTGCGAGAGCGCGCGGAGAGTACCTCTCG 1888
2Y 2027 CGCGCAAGCGCGCTTCTGATCTGCTGCTGCTGCGAGAGCTGCGGAGAGCGCGCGCG 2086
Db 1889 CGCGCAAGCGCGCTTCTGATCTGCTGCTGCTGCGAGAGCTGCGGAGAGCGCGCGCG 1948
2Y 2087 GCGACGCTGCTTCTGCGGCTGCGAGAGAGAGTACGAGTCTGCGGAGTCTTACG 2146
Db 1949 GAGACGCTTCTTCTGCGGAGAGAGAGTACGAGTCTTCTGCGGAGTCTTACG 2008
2Y 2147 AGGCGATCAAGAGCGCGCATCAACGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2198
Db 2009 AGGCGATCAAGTCTGCGGAGAGTCAACGAGTCTTCTGCTGCTGCTGCTGCTGCTGCT 2060

RESULT 9
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koester, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC

CURRENT APPLICATION NUMBER: US/09/624, 693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 47.0%; Score 1136.6; DB 4; Length 2787;
Best Local Similarity 71.6%; Pred. No. 1.6e-21;
Matches 1557; Conservative 1; Mismatches 590; Indels 27; Gaps 4;

2Y 27 AGCACTGGCATGGGCGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 86
Db 636 AGCTCTCAAGTGGGCGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 695
2Y 87 CAACGGCTGCGACCGCGCTGCGACCAAGTGGCTGCGGCGCCCACTTGGCTTCCGCGCG 146
Db 696 GAACGGGTTGACCG 755
2Y 147 CAGCGCGGCTGATGATGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 206
Db 756 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815
2Y 207 CCTGAGCGACCCCGACCGACGAGTCTGAGCTCAAGCGGCTGACGCTCAACGCTGCTGTA 266
Db 816 GCGCGAGCG 875
2Y 267 GCTTGTGCG 326
Db 876 GCTGCTGCG 935
2Y 327 CGGAGCGGCTGCAAGAGCGCTGCACTTCTCTCAAGCGCGCGCGCGCGCGCGCGCGCG 386
Db 936 CGAAGAGTCAATGCGAGTGTGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 995
2Y 387 AGTCAACAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
Db 996 TGTCAAGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
2Y 447 GAGGCGCTCATGAGAGCGAGCTCTGCGGCTGAGAGCGCGCGCGCGCGCGCGCGCG 506
Db 1056 AAGGCGCTGCTGAGAGCGAGCTCTGCGGCTGAGAGCGCGCGCGCGCGCGCGCGCG 1115
2Y 507 CGTGGAGCGGCGCTGAGAGCACGCTTCTGCTGAGAGTCTGCTGCGCGCGCGCATGTAT 566
Db 1116 GCTGCTGCGGCGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
2Y 567 CGGCTCAATGCTGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 626
Db 1176 CGGCTCAATGCTGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
2Y 627 CACCACTTCTTGAACCAAGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
Db 1236 CACCACTTCTTGAACCAAGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
2Y 687 GTGCGGCACTTCAAGCGCGCTTCTGATCATGCGCGCGCGCATCAACCGGTCAACCGGAGT 746
Db 1296 GTGCGGCACTTCTTCCCGCTCTTATCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
2Y 747 CAAGTTCAGTTTTCAGAGGAGGAGGAGTATGTTTGGCGCGCGCGCATCTC 806
Db 1356 GAAGTTCAGT-----CGACGCAAGTATGTTTCCCGCGCGCGCGCGCGCGCGCG 1403
2Y 807 GCTCTTGGCTGAGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
Db 1404 GCTCAAGGCTTTCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463

QY 867 AAGGCGCTCTCCGCTCGATGGGCAACCTCACTGTCAGACGATGCGACATGCTCTGCT 926
DB 1464 CAGGCGCTCTCCGCTCGATGGGCAACCTCACTGTCAGACGATGCGACATGCTCTGCT 1523
QY 927 CTTCTGCAAGGCTTTGACGCTCTCAAGGTCAGGTCATGTCGAGCAGAGGCTCTGCT 986
DB 1524 CTTGCAAGGCTCTCAAGGTCAGGTCATGTCGAGCAGAGGCTCTGCT 1583
QY 987 GCGCGCTTATCCAGAGCTGTCGCGCGCGACCGCGCGAGGTCGAGGTCGCGCA 1046
DB 1584 CACCATTCCTCAGAGCTGTCGCGCGCGACCGCGCGAGGTCGAGGTCGCGCA 1643
QY 1047 CATCCGACGCTCTTCCGCTCGCTGCTTTCGCTGTCAGGTCAGGTCAGGTCAGGTC 1106
DB 1644 CATCCGACGCTCTTCCGCTCGCTGCTTTCGCTGTCAGGTCAGGTCAGGTCAGGTC 1703
QY 1107 CAGGACGAGAGGAGGATTTCTGCGGAGACCGCTACCCGCTCGGACGTCGCTCAATT 1166
DB 1704 CAGGACGAGAGGAGGATTTCTGCGGAGACCGCTACCCGCTCGGACGTCGCTCAATT 1763
QY 1167 CTTGCGCGCTCTGTCAGGAGATGATGAGGCTTCACTGCTCTGCTCGAG---AA 1223
DB 1764 GCTCGCTCTCTGTCAGGAGATGATGAGGCTTCACTGCTCTGCTCGAGGCTG 1823
QY 1224 CAAACGAGAGAGGACCGCTCTGCTGCTGAGAACAGAGACCGGCGACGCGG 1283
DB 1824 TCAGTGCAGACCGAGAACCGCTGATGATGAGAACAGAGAACCGACATGCGG 1883
QY 1284 CAACCTTCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1343
DB 1884 AGCTTCATGAGGAGAGGTCGAGAACACATGAGAGAGCTGCTGCTGCTGCTGCT 1943
QY 1344 CATGCGACAGCTCACTTCAAGGTCAGGAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1403
DB 1944 GATGCGACAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 2003
QY 1404 CTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463
DB 2004 CTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063
QY 1464 CATTCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1523
DB 2064 CATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2123
QY 1524 GCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1583
DB 2124 GCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2183
QY 1584 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1643
DB 2184 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2243
QY 1644 GCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1703
DB 2244 CTTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2303
QY 1704 TCTCCGACTCTCTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1763
DB 2304 GCTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2360
QY 1764 CGAGGTCAGAGGCTGCTCAACAGGTCGAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 1823
DB 2361 CAGGTCAGAGGCTGCTCAACAGGTCGAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 2420
QY 1824 GCGCTGCGAGAGGCTTCTGTCAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 1883
DB 2421 GCGCTGCGAGAGGCTTCTGTCAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 2480
QY 1884 GCGCTGCGAGAGGCTTCTGTCAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTC 1943
DB 2481 G-----GAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2531
QY 1944 GCGCATCTGCTCAAGGCTGCGCAACGCTTCTGCGAGACCGCTCTTTCGAGGCT 2003

DB 2532 GGTATTCGCTGCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2591
QY 2004 GCGGCGGCAAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063
DB 2592 GCGGCGGTCAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2651
QY 2064 GCTGCGGTCAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2123
DB 2652 GCTGCGGTCAGGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2711
QY 2124 GAGCAAGCTTCGCGCATCTGAGGTCATGAGGTCATGAGGTCATGAGGTCATGAGGTCATGAGGTC 2183
DB 2712 CACCAAGCTTCAGGTCATCTGAGGTCATGAGGTCATGAGGTCATGAGGTCATGAGGTCATGAGGTC 2771
QY 2184 CAGATGCTGCGGTA 2198
DB 2772 CAGATGATGGATA 2786

RESULT 10
US-09-624-693A-14
/ Sequence 14, Application US/09624693A
/ Patent No. 6355468
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ FILE REFERENCE: 29479/500NSC
/ CURRENT APPLICATION NUMBER: US/09/624,693A
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 14
/ LENGTH: 2311
/ TYPE: DNA
/ ORGANISM: Amanita muscaria
/ FEATURES:
/ NAME/KEY: CDS
/ LOCATION: (18)..(2237)
US-09-624-693A-14

Query Match 12.5%; Score 302.8; DB 4; Length 2311;
Best Local Similarity 54.2%; Pred. No. 4.5e-50;
Matches 719; Conservative 1; Mismatches 573; Indels 33; Gaps 4;

Db ATTGCAACCCATCAATCAAGTATATACAGGTCATCAAAAGCCGGAATTCGCCAAATT 791
732 ATTTGGCGCGAGGCGCATCTCGCTTTTGTGCTGAGGCACTGCTCTCGGCCGAAG 843
784 ATGTTGGCGCGAGGCGCATCTCGCTTTTGTGCTGAGGCACTGCTCTCGGCCGAAG 843
Db GGATCTCGAAGATGCTTGGCTCTGCAATATCAACCTTCCACATGGAATGAAA 851
792 GGATCTCGAAGATGCTTGGCTCTGCAATATCAACCTTCCACATGGAATGAAA 851
2y GAGGCTCTCGGCTGCTGCAACGGAACGGCGCTCCGCGCTCGATGGGACCCCTCACTCG 903
844 GAGGCTCTCGGCTGCTGCAACGGAACGGCGCTCCGCGCTCGATGGGACCCCTCACTCG 903
Db GAACCTCTGGATTTTGAATGGAGCGCATTTCCGCAATCTGGGCACTTTAGGCTTA 911
852 GAACCTCTGGATTTTGAATGGAGCGCATTTCCGCAATCTGGGCACTTTAGGCTTA 911
2y CACGACTCGCATGCTCTCGCTCTCTCGAGGCTTGAACGGCTCTCAAGTGAGGCC 963
904 CACGACTCGCATGCTCTCGCTCTCTCGAGGCTTGAACGGCTCTCAAGTGAGGCC 963
Db AACGAGCTATCCATCTGCTTGTGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGA 971
912 AACGAGCTATCCATCTGCTTGTGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGA 971
2y ATGTCGCGCCACAGAGGCTGCTTGGCGCGCTTCACTCCAGACGCTGCGCGCGCGCGCG 1023
964 ATGTCGCGCCACAGAGGCTGCTTGGCGCGCTTCACTCCAGACGCTGCGCGCGCGCG 1023
Db TTGATAGGCACTCGGCTTCTCATGCAACCGCTTCAATATCCACCGCAACCAATCC 1031
972 TTGATAGGCACTCGGCTTCTCATGCAACCGCTTCAATATCCACCGCAACCAATCC 1031
2y GCGCAGTCAAGTTCGCGCAACATCCGCAACGCTCTTCCGCGCTGCTGCTTTCGCTT 1083
1024 GCGCAGTCAAGTTCGCGCAACATCCGCAACGCTCTTCCGCGCTGCTGCTTTCGCTT 1083
Db GGTCAAGTGAATGCTGCTGAGAACATTTGGATTTGCTGAGTGAATTAATTGGCTCAG 1091
1032 GGTCAAGTGAATGCTGCTGAGAACATTTGGATTTGCTGAGTGAATTAATTGGCTCAG 1091
2y GAGCAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1143
1084 GAGCAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1143
Db TTAGAAGAGCAAGATTCGCTTGAAGACGATTAATACCCCTTCCGAGAGAGAGAG 1151
1092 TTAGAAGAGCAAGATTCGCTTGAAGACGATTAATACCCCTTCCGAGAGAGAGAG 1151
2y CCGCTCCGCAAGTTCGCTTCAATCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1203
1144 CCGCTCCGCAAGTTCGCTTCAATCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1203
Db CCACTCCGCAATTCGCTTCAATCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1211
1152 CCACTCCGCAATTCGCTTCAATCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1211
2y TCGACTCTCTGCTGAG-----AACAAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
1204 TCGACTCTCTGCTGAG-----AACAAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db CAGACTGTAACGAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1271
1212 CAGACTGTAACGAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1271
2y GAGAAACAAG 1317
1258 GAGAAACAAG 1317
Db GAGACTGAGATTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1331
1272 GAGACTGAGATTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1331
2y GAGAAGACAG 1377
1318 GAGAAGACAG 1377
Db GAGAAAGACAG 1391
1332 GAGAAAGACAG 1391
2y TTGCTCAACGCTGCTGATGAACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1437
1378 TTGCTCAACGCTGCTGATGAACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1437
Db TTAGTCAATCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1451
1392 TTAGTCAATCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1451
2y CTCACATATCAAG 1497
1438 CTCACATATCAAG 1497
Db CTCACATATCAAG 1504
1452 CTCACATATCAAG 1504
2y CACCTTGGCAACCGAGTCACTACCTTGTCAAGCCGCAAGATGGGCAACAGAGCGCTC 1557
1498 CACCTTGGCAACCGAGTCACTACCTTGTCAAGCCGCAAGATGGGCAACAGAGCGCTC 1557
Db AGCAGACTCTGAGCG 1559
1505 AGCAGACTCTGAGCG 1559
2y AACTGCTGCTGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1617
1558 AACTGCTGCTGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1617
Db AACTGCTGCTGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1619
1560 AACTGCTGCTGCTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1619
2y CTCTGCGCGCGAG 1677
1618 CTCTGCGCGCGAG 1677
Db CTGATGCGCTCTTACTGTTATATCTGATGCGCAAGTCTGAGCTTCGAGCGCTTCAAG 1679
1620 CTGATGCGCTCTTACTGTTATATCTGATGCGCAAGTCTGAGCTTCGAGCGCTTCAAG 1679
2y GACTTC 1683
1678 GACTTC 1683
Db GAGTTC 1685
1680 GAGTTC 1685

APPLICANT: Bloksberg, Leonard N.
APPLICANT: Hayakkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
TITLE OF INVENTION: Modification of Plant Lipid Content
FILE REFERENCE: 11000.10034U
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 245
LENGTH: 1455
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-245
Query Match 8.6%; Score 207.2; DB 4; Length 1455;
Best Local Similarity 52.4%; Pred. No. 1,4e-31;
Matches 623; Conservative 0; Mismatches 528; Indels 39; Gaps 6;
184 CAGCTCGAGATGCTGCGAG 243
299 CAGCTCGAG 358
244 GGGTACAG 303
359 GGGAG 415
304 GTCCAG 353
416 GTCCAG 475
364 CAGCTTCAG 417
476 AGCATGAACAG 535
418 AGCAG 477
536 GCGAG 594
478 GTGAG 537
595 -----GATCTTCGAG 634
538 CTGAG 597
635 CAATCTCAACAG 694
598 GCGAG 657
695 GGCATTCGATTTGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 754
658 ATCTGCGCGCGAG 717
755 TGCCTGCGCGCGAG 814
718 GCGAG 777
815 GCGAG 837
778 AGATGATGATTTGCGCGAG 837
872 CTGAG 931
838 CCGAAG 897
932 CCAAG 991

RESULT 11
US-09-615-192A-245
; Sequence 245, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:

TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
J5-09-252-991A-16475

Query Match 5.2%; Score 126.8; DB 4; Length 1548;
Best Local Similarity 47.1%; Pred. No. 5.1e-16;
Matches 571; Conservative 0; Mismatches 592; Indels 49; Gaps 4;

2Y 516 CGGCTCGAGAACACGCTTCGCTCGAGGTGTCGCGCGCGCATGTCATCCGCGTCAA 575
DB 351 CGGCGTCGAGCGCGCTGACGACGCCATGTGCGGCTGTGTATGCTCAAGTGA 410
2Y 576 CTCGCTACGCGTGGACCACTGCGCGCTGCTGCTGCTTGAAGCGCTCAACTT 635
DB 411 CAGCTGCGCGCGGCTTCTCGGCATCCGCGCAAGTATGACGCGCTGATCCGCT 470
2Y 636 CTTGAACCAACGCAATCAGCCCATGTCCTCCGCGCTTCATCTCGCGCTCGGCGCA 695
DB 471 GATCAACCGCCGAGGTATCCGACATCCGCTGAAGGCTCGGTGGTCCGCGCA 530
2Y 696 CCTACGCGCGCTCTGTCATCGCGCGGCGCATCAGCGCTCACCGCGATCAAGGTTCA 755
DB 531 CTTGCGCGCTGCGCGCAATGTGCTGTCTGATGCGGAAAGCGCGCGCGCATCG 590
2Y 756 CGTTTTCGACGAGGGAACGAAAGATCATGTTTGGCGCGGAGCCATCTCGCTTTTG 815
DB 591 CGGTG-----AGTGTGCGCGCGCGCGGAGCGCTGCGGTGCGCG 632
2Y 816 TCTCGAGGCACTGCTGCTCGCGCGCGGAGGAGGTCTGGTCTGTCAGAGAAAGCGCGT 875
DB 633 GCTGAGGCGCTGACCTTGCGCGCGGAGAAAGGCTTGCTGCTCAATGACCCAGT 692
2Y 876 CTCGCGCTGATGAGGCACTCACTGCTGACAGATCCGACATGCTCTGCTCTTCTGCA 935
DB 693 GTCCACCGCTACGCGCTGCGCGGCTTGTGAGCGCGAGACTGTTCGCGCGCGCAC 752
2Y 936 GGCCTTGAAGGCTCTACGCTGAGGAGGCTGTGCGCGAGAGGCTGCTTCCGCGCTT 995
DB 753 GGTCTGCGCGCGGCTCAGCGTCAAGGCTGCTGCTGCGCGCGCGCTTCTGATGCGCG 812
2Y 996 CATTCACGACGCTGCGCGCGCGCGCACCGCGCGAGTGCAGGTCGCGCAATCCGCA 1055
DB 813 CATTCATGCGCGCGCGCGCGCGAGCTGAGGAGATGACATGCGCGCGCTTATGCGCACT 872
2Y 1056 GCTCTCTTCCGCGCTGCTGCTTGTGCGTTGACGACGAGAGAGGTCAAGTCAAGACGA 1115
DB 873 GGTCAACCGCAAGACGAGGTGGCGCGCTCCCATGAGAGTGCACAAGT----- 923
2Y 1116 CGAGGCAATTCTTCCCGCAAGACGCTACCGCTCCGACGCTGCTTCAATTCTCGGCGC 1175
DB 924 -----CGAGAACCTTATTCGCTGCGTTGCCAGCGCGAGGTGAGGCGCG 968
2Y 1176 GCTCGTGAAGACATGATGACAGCTTACTCTCTGCTCGAGAAACAAGACGAC 1235
DB 965 CTGCTGACCCAGATGCGCAGAGCGCGCGAGGTGTGAGATGCAACCAAGCGGTGTC 1028
2Y 1236 CGACAAACCGCTCTCTGACGTGTGAGAACAGACGCGCGCAAGCGCGCACTTCCAGCG 1295
DB 1029 CGACAAACCGCTGTATTGCGCGCGAGGCGAGGTGATCTCCGCGCGCAACTTCCACG 1088
2Y 1296 GTTCGCTGTCTGATTTGATGAGAAAGACAGGTGCACTGCGCGCTCATTCGCAAGCT 1355
DB 1089 CGAACCGGTGAGATGCGCGCGCGCAACCTTGCGCTGCGGTGCGAGATCGTTCG- 1147
2Y 1356 CAACCTTACGACGATGACCGAGATTGCTCAAGCTGCGCAAGAACCGCGCGCTTGTG 1415
DB 1148 -----TGTGGAAGCGCGCATCTGCTGATGATGACATGACATGTGAGAGTTGCGCGC 1202
2Y 1416 CTTGCTGCGCGAGACCGCTGCTCAATATACGCGCAAGGCTTGAACATTCATGCTG 1475
DB 1203 GTTCTGTGTGAGCAAGCGCGGAGTCAACTTCGCGCTTCAATGATGCGCGCAAGTCAAGCGCG 1262
2Y 1476 TGTCTTACGCTTGGAGCTGCGCGACCTTGCACACCGGATCACTTCTGTCAGCGCGC 1535

DB 1263 CGGCTGCGCGAGCAACAAAGGCGCTGCGCGCATCCGCGCAGGTGACAGCTGCGCAG 1322
QY 1536 AGAATGGAACACAGGCGCTCAACTGCTGCTCATCTCGCGCGCGCGCACTGCCA 1595
DB 1323 CTCGCGCAACAGAGAACCATGTGTGATGCGCGCAAGCGCGCAAGCGCTCGGCG 1382
QY 1596 GAGCAACGAGCTCTTCTGCTGCTGCGCTGCACTGATCACTGACGCTTCCGCGACT 1654
DB 1383 GATGCGCGAGACGTCGCGCGGAGTTCGCGGTGAGTGTGCGCGCGCGCGCGCGCG 1442
QY 1655 TCGACCTTCGCGCGATGAGCTGACCTTCAAGAACAGATTCGACCGCGCTTCCGCGACT 1714
DB 1443 TGAAGCTTCGCGAGGCGCTGAGAGATTCGCGCAAGCTGAGAGCGCTCGCGCTGCTG 1502
QY 1715 TCTTCAGAGAC 1726
DB 1503 GCGCAAGGTTC 1514

RESULT 14
US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16126
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

Query Match 5.2%; Score 126.8; DB 4; Length 2295;
Best Local Similarity 47.1%; Pred. No. 5.4e-16;
Matches 571; Conservative 0; Mismatches 592; Indels 49; Gaps 4;

QY 516 CGGCTCGAGAACACGCTTCGCTCGAGGTGTCGCGCGCGCATGTCATCCGCGTCAA 575
DB 1168 CGGCGTCGAGCGCGCTGACGACGCCATGTGCGGCTGTGCTCAAGTGA 1109
QY 576 CTGCTCACGCGTGGCACTGCGCGCTGCTGCTGCTTGAAGCGCTCAACAACTT 635
DB 1108 CAGCTGCGCGCGGCTTCTCGGCATCCGCGCAAGTATGACGCGCTGATCGGCT 1049
QY 636 CTGGAACACCGCATCAAGCCATGCTCCCTTCGCGCGCTCATCTCGCGTGGCGGA 695
DB 1048 GATCAACCGCAAGGTATTCGCAATCCGCTGAAGAGGTGCGGTGCTGCTCGCGCA 989
QY 696 CTTGACCGCTCTCTGTCATGCGCGCGCGCATACCGGTGACCCCGAGCTCAAGTTCA 755
DB 988 CTTGCGCGCGCTGCGCGCAATGTGCTGTGCTGATGCGGAAAGCGCGCGCGCATCG 929
QY 756 CGTTTTCGACGAGGGAACGAAAGATCATGTTTGGCGCGAGGCACTTCGCTTTTG 815
DB 928 CGGTG-----AGTGTGCGCGCGCGCGGAGCGCTGCGGTGCGCG 887
QY 816 TCTCGAGGCACTGCTGCTGCGCGCGAAGAGGATCTCGGTGTGTCACGAAAGCGCGCT 875
DB 886 GCTGAGCGCGCTGACCTTGGCGCGGAGAGAGGCTGCGCTGCTCAATGACCCAGGT 827
QY 876 CTCGCGCTGATGAGGACCTTCACTGTCAGACTGTCGACATGTGCTGCTCTTCTGCA 935
DB 826 GTCCACCGCTTACGCGCTGCGCGGCTGTGAGGCGAGAGACTGTTCGCGCGCGCAC 767

QY 936 GGCCTTGACGGCTCTTCACGGTGAAGCCATAGTGGCCAGAGGGCTGCTTGGGCGGTT 995
 Db 766 CGTCTCGCGGGCTAGAGCTGAGGCCATGCTTCGGTTCCGGGGCGCCGTTGATGCGCG 707
 QY 996 CATTCACGAGCTCTGCGCGCCGACCCCGCCAGGTGAGGTGCGCGCAACATCCGAC 1055
 Db 706 CATTCATGCGCGCGCGCGCCAGCGTGGGATGACATGAGCTGGCGCGGCTTATCCGACT 647
 QY 1056 GCTCCTTTCCGGGTCTGCTTGGCGGTGAGACAGAGAGAGGTCAAGGTCAAGAGACA 1115
 Db 646 GCTCACCAGAGAGAGAGGTGCGCGCTCCCATAGAGGTGCAAGAGT----- 356
 QY 1116 CGAGGACATTTCCGACGAGACCGGCTACCGGCTCCGACGTGCGCTAGTTCCCTGGGCG 1175
 Db 595 -----CGAGACCCCTATTGCTGGTGGCAGCGCGAGGTATGGGCGC 551
 QY 1176 GCTCTGAGAGACATGATGACAGCTTACTCTGCTCTGCTCGAAGAACACAGACGAC 1235
 Db 550 CTGCTGACCCAGATGCGCCAGCGCGAGGTGAGATGAGAGCAACGCGGTGTC 491
 QY 1236 CGACAAACCCGCTCTCTGACAGTGAAGAACAGACCGCGCACTTCCAGCG 1295
 Db 490 CGACAAACCCGCTGATTTGCGCGCGAGGCGACGTGATCTCCGCGCACTTCCAGCG 431
 QY 1296 GTGCGGTGCTGATTTGATGAGAGAACAGAGGTGCACTGCGCTCAATCGCAGCT 1355
 Db 430 CGAACCGGTGCGATGCGCGCGACACCTGCGCTGCGTGGCCGAGATCGGTTGC- 372
 QY 1356 CAACCTTCAGCAGTGCACCGAGTTGCTCAAGCTGCCATGAAACCGCGCGCTGCTTGTG 1415
 Db 371 -----TGTGGAACCGCGCATCTGCTATGATGAGACATGCAATGTCGAGTTGCCGC 317
 QY 1416 CTTGCTGCGCAGAGACCCGCTGCTCACTATCAGCGAGCGGCTTGGACATTCACATGCG 1475
 Db 316 GTTCTGAGTGGCCAAACGCGGGGTCACTCGCGCTTCAATGCCCGAGGTCAACCGCGC 257
 QY 1476 TGTTCATGCTTGGAGCTGCGCACCTTGCACACCGGTCATACCTTGTCCAGCGCGC 1535
 Db 256 CCGCGTGGCAGCGCAACAGGCGCTGGCCATCGCGCAGCGCTGCAGCGCTGCGCAGC 197
 QY 1536 AAGATGAGGCAACACGAGCGCTCAACTGCTGCTCTGATCTCCGCGCGCGCACTGCCA 1595
 Db 196 CTGCGCAACACGAGAGACATGCTGATGATGATGATGATGATGATGATGATGATGATG 137
 QY 1596 GGCACACGAGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1654
 Db 136 GATGCGCAGAGAGCTCCCGCGGATCTGCGCGTGGAGTGGGCTGGCGCGCTGCGCAGG 77
 QY 1655 TCGACCTCGCGCGCATGAGCTGCACTTCAAGAGCATTCGACCGCGCTTCTCCGACTG 1714
 Db 76 TCGACTTCGCGAGGGCTTGAAGAGTTGCGGAGGTGAGAGAGGTGCGCGCTGCTGCTG 17
 QY 1715 TCTTCACAGCAG 1726
 Db 16 GCGACAGGTGC 5

RESULT 15
 US-09-489-039A-6404
 ; Sequence 6404, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 6404

LENGTH: 1527
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-6404
 Query Match 4.5%; Score 108.4; DB 4; Length 1527;
 Best Local Similarity 48.2%; Pred. No. 1.9e-12;
 Matches 405; Conservative 0; Mismatches 396; Indels 39; Gaps 2;
 QY 517 GGCCTGAGAACACGCTTCGCTGAGGTGCTGCGCGCGCCAGGTATCGCGTCAAC 576
 Db 259 GGCCTTGGCGAGCGCGCTGACGATGATGCGCGCGCTGATATGATGATGATTAAC 318
 QY 577 TCGCTCAAGCTGACACTGCGCGCGCGCTGCTGCTGCTTGAAGCGCTCAACATTC 636
 Db 319 AGCTGTCCCGGTTCTTCGGGATCGGCTTGAAGGTATCCAGCGCTGATGGCTG 378
 QY 637 TTGAACACCGCATACAGCCCATGCTCCCGCTCGCGCTCATCTCGCGTGGCGAC 696
 Db 379 GTGATGCGCGGTGATGACCGCGTGAATCCGCGCAAGAGCTCGCTCGCGCGAC 438
 QY 697 CTCAGCGCGCTCTGTCATGCGCGCGCGCATACCGGTGACCCGACGTCAAGTTAC 756
 Db 439 CTGCGCGCGCTGCGCATATGCTGACCTGATCGGAGAGGTAAAGCGCGGTACCC 498
 QY 757 GTTTGACAGAGGAAACGAGAAATCATGTTTGGCGCGAGGCACTCGCTTGTG 816
 Db 499 GCGCGCACTGCG-----TCCGCGCACCGAAGCGTGGTCAAGTGGG 543
 QY 817 CTCAGGAGTGTGCTCTGCGCGCGAGAGAGGTGCTGCTGCTGCTCAACCGCGCTG 876
 Db 544 CTGAGCGCATATACCTGCGCGCGAAGAGGCGCTGCGCTACTCAACGCGCACCGCGC 603
 QY 877 TCGCGCTGATGAGGACCTCAGTCTGACAGCATGCGACATGCTCGCTGCTGCGAG 936
 Db 604 TCGACCGCTTTGCTCTGCGCGCGGTGTTGAAAGCGAGAGATGTTGCGCTCCGCGTC 663
 QY 937 GCTTGAAGCTCTCAAGGTGAGAGCCATGCTGCGCGCAGAGGCTGTTGCGCGCTTC 996
 Db 664 GTCTGCGGGGCTTAACACCGAAGCGCGCTGCTGCTCAAGTGTGATGATGCGCGC 723
 QY 997 ATCCAGACGTCTGCGCGCGCGCACCCCGCGCAGGTGAGGTGCGCGCACATCCGACG 1056
 Db 724 ATCCATGATGCGCGCGCGCACCGCGCGCAGATGATGCGCGCGCTGATGCCACCTG 783
 QY 1057 CTGCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1116
 Db 784 CTGACCGACATGACCGGATCTGCGAGTGCACATTAATGACAGAGGT----- 833
 QY 1117 GAGGGCATTTTGGCGACAGACCGCTACCGCTTCCGACGTGCTGCTGCTGCTGCTG 1176
 Db 834 -----GAGAGACCGGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 879
 QY 1177 CTGCTGAGGACATGATGACAGCGCTACTGACATCTGCTGCTGAGAACACAGACGACC 1236
 Db 880 TGCCTGACGAGATCCGCCAGCGGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939
 QY 1237 GACAAACCGCTCTCTGACAGTGCAGAACAGACAGCGCGACGCGCGCACTTCCAGCG 1296
 Db 940 GATTAACCGCTGCTGCTTGGCGCGGAAATGACGTCATCTCGGGGATTAATCTCATGCC 999
 QY 1297 TCGGCTGTCTGATTTGATGAGAAAGACAGGCTGCGACTGCGCTCATCGGCAAGTCT 1356
 Db 1000 GAAACCGGTGCGATGCGCGCGGATATATGCGGCTGCGGATGCGGAAATGCGGCTGCT 1059

Search completed: September 12, 2004, 04:26:43
 Job time : 190.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M nucleic - nucleic search, using sw model

Run on: September 11, 2004, 23:12:54 ; Search time 1570.32 Seconds
(without alignments)
7750.794 Million cell updates/sec

Title: US-09-939-408A-12

Perfect score: 2419
1 ctctcgtctactactaacc.....gagttcaaaaaaaaaa 2419

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10C_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US00_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2418.6	100.0	2419	9	US-09-939-408A-12
2	1826.6	75.5	2741	9	US-09-939-408A-28
3	1704.4	70.4	2163	9	US-09-939-408A-29
4	1629	67.3	2475	9	US-09-939-408A-20
5	1196.6	49.5	2151	9	US-09-765-873A-9
6	1196.6	49.5	2151	15	US-10-138-970A-23
7	1196.6	49.5	2151	15	US-10-188-523B-9
8	1196.6	49.5	2151	16	US-10-374-366-3
9	1196.6	49.5	2439	9	US-09-939-408A-18
10	1195	49.4	2151	9	US-09-765-873A-7
11	1195	49.4	2151	15	US-10-188-523B-7
12	1195	49.4	2151	16	US-10-374-366-1
13	1195	49.4	2151	16	US-10-374-366-9
14	1195	49.4	2151	16	US-10-374-366-13

15	1195	49.4	2151	16	US-10-374-366-23	Sequence 23, Appl
16	1193.4	49.3	2151	16	US-10-374-366-5	Sequence 5, Appl
17	1193.4	49.3	2151	16	US-10-374-366-7	Sequence 7, Appl
18	1193.4	49.3	2151	16	US-10-374-366-15	Sequence 15, Appl
19	1193.4	49.3	2151	16	US-10-374-366-21	Sequence 21, Appl
20	1191.8	49.2	2151	16	US-10-374-366-17	Sequence 17, Appl
21	1190.2	49.2	2151	16	US-10-374-366-11	Sequence 11, Appl
22	1190.2	49.2	2151	16	US-10-374-366-19	Sequence 19, Appl
23	1188.8	49.1	2061	9	US-09-765-873A-31	Sequence 31, Appl
24	1188.8	49.1	2061	15	US-10-188-523B-31	Sequence 31, Appl
25	1186.8	47.0	2787	9	US-09-939-408A-16	Sequence 16, Appl
26	1136.6	43.2	2703	16	US-10-439-478-1	Sequence 1, Appl
27	802.4	33.2	2703	16	US-10-439-478-24	Sequence 24, Appl
28	696.8	28.8	2465	15	US-10-138-970A-3	Sequence 3, Appl
29	545	22.5	1020	15	US-10-138-970A-3	Sequence 15, A
30	424.8	17.6	2554	17	US-10-767-701-15815	Sequence 15815, A
31	423.2	17.5	2539	13	US-10-425-114-26376	Sequence 26376, A
32	408.2	16.9	2325	17	US-10-767-701-15813	Sequence 15813, A
33	402	16.6	2544	17	US-10-437-963-55261	Sequence 55261, A
34	401.6	16.6	2454	13	US-10-425-114-34124	Sequence 34124, A
35	400.2	16.5	2417	13	US-10-425-114-31197	Sequence 31197, A
36	400.2	16.5	2468	13	US-10-425-114-30888	Sequence 30888, A
37	400.2	16.5	2469	13	US-10-425-114-31903	Sequence 31903, A
38	385.8	15.9	2972	17	US-10-437-963-48311	Sequence 48311, A
39	383	15.8	2774	17	US-10-437-963-38582	Sequence 38582, A
40	382.8	15.8	2774	17	US-10-437-963-41428	Sequence 41428, A
41	382.2	15.8	2079	13	US-10-425-114-1334	Sequence 1334, Ap
42	380.4	15.7	1092	17	US-10-439-479-9	Sequence 9, Appl
43	376.2	15.6	2340	13	US-10-425-114-2857	Sequence 2857, Ap
44	350.8	14.5	2145	17	US-10-437-963-40094	Sequence 40094, A
45	346	14.3	2535	17	US-10-437-963-36111	Sequence 36111, A

ALIGNMENTS

RESULT 1
US-09-939-408A-12
Sequence 12, Application US/09939408A
Patent No. US20020102712A1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
APPLICANT: Koestera, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939, 408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0.
SEQ ID NO 12
LENGTH: 2419
TYPE: DNA
ORGANISM: Rhodotomula graminis
FEATURE:
NAME/KEY: CDS
LOCATION: (37)..(2196)
NAME/KEY: modified_base
LOCATION: (494)
OTHER INFORMATION: Other information: Y = t or c
NAME/KEY: unsure
LOCATION: (493)..(495)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408A-12

Query Match 100.0%; Score 2418.6; DB 9; Length 2419;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTCCGCGCTCACTCTAAACCCGGTCACTAGACCTTGGCATGGCCCCCTTCTTGGACTCGCTC	60
Db	1	CTCTGCGCTCACTCTAAACCCGGTCACTAGACCTTGGCATGGCCCCCTTCTTGGACTCGCTC	60
QY	61	GCCACCAAGCTCGCCAAACGGCTTTACCAACGGCTCGACGGCGCTCGACCAAGTCGGCT	120
Db	61	GCCACCAAGCTCGCCAAACGGCTTTACCAACGGCTCGACGGCGCTCGACCAAGTCGGCT	120
QY	121	GCGGACCCCACTTGGGCTCTCCGCGCACGGCCCGCTTGGATGGCCACGCCGACAG	180
Db	121	GCGGACCCCACTTGGGCTCTCCGCGCACGGCCCGCTTGGATGGCCACGCCGACAG	180
QY	181	TGCGACGCTCGAGATCGTGCAGAGAGCTCTCAGAGACCCCAACCGAGAGCTGTGAGACTC	240
Db	181	TGCGACGCTCGAGATCGTGCAGAGAGCTCTCAGAGACCCCAACCGAGAGCTGTGAGACTC	240
QY	241	AGCGGGTACAGCCTCACCGTCCGTGAAGTTGTGCGCGCCCGCAAGAGGCGCAGGGTTC	300
Db	241	AGCGGGTACAGCCTCACCGTCCGTGAAGTTGTGCGCGCCCGCAAGAGGCGCAGGGTTC	300
QY	301	GCGGTCCAGAAACGACGACGAGATCCCGCGACCGCTGCACAGAGCTTGAATTTCTTAAG	360
Db	301	GCGGTCCAGAAACGACGACGAGATCCCGCGACCGCTGCACAGAGCTTGAATTTCTTAAG	360
QY	361	GCCCAAGCTTGAAACTGGCTTACGGAATCACAGGGTTTCGGTGGCTCGGCGACAGC	420
Db	361	GCCCAAGCTTGAAACTGGCTTACGGAATCACAGGGTTTCGGTGGCTCGGCGACAGC	420
QY	421	AGACCTAAGATGACGTCAAGCTTCCAGAAAGCGCTCATCGACACCAAGCTTGGGGGTG	480
Db	421	AGACCTAAGATGACGTCAAGCTTCCAGAAAGCGCTCATCGACACCAAGCTTGGGGGTG	480
QY	481	ACGCGCAAGTCCGCTCTTCAAGCGTCGGAACCGGCGCTTCGAGAACAGCTTCCGCTC	540
Db	481	ACGCGCAAGTCCGCTCTTCAAGCGTCGGAACCGGCGCTTCGAGAACAGCTTCCGCTC	540
QY	541	GAGGTGCTCGCGGCGCCATGATCATCCGCTCAACTGGCTCAAGCTGGCGCACTGGCC	600
Db	541	GAGGTGCTCGCGGCGCCATGATCATCCGCTCAACTGGCTCAAGCTGGCGCACTGGCC	600
QY	601	GTCGCGCTCGTCTCTTTAAGCGCTCACCAACTTTTGAACACCGCATACGCCCATC	660
Db	601	GTCGCGCTCGTCTCTTTAAGCGCTCACCAACTTTTGAACACCGCATACGCCCATC	660
QY	661	GTCCCCCTCGCGGCTCCATCTCGGCGTGGGCGACCTCAGCCCGCTCTGTACATCGCC	720
Db	661	GTCCCCCTCGCGGCTCCATCTCGGCGTGGGCGACCTCAGCCCGCTCTGTACATCGCC	720
QY	721	GCGGCCATCACCGGTCAACCCGAGCTCAAGTTCAAGTTTGCACGAGGAAACGAGAG	780
Db	721	GCGGCCATCACCGGTCAACCCGAGCTCAAGTTCAAGTTTGCACGAGGAAACGAGAG	780
QY	781	ATCATGTTTTCGCGGACGAGCATTCGCTCTTGGTTCAGAGGAGTCTCGCTCGGCCG	840
Db	781	ATCATGTTTTCGCGGACGAGCATTCGCTCTTGGTTCAGAGGAGTCTCGCTCGGCCG	840
QY	841	AAGAGAGGCTCTCGGTCTGCTCAACGAAACGACCGCTTCGCTCGATGAGCAACCTCAAT	900
Db	841	AAGAGAGGCTCTCGGTCTGCTCAACGAAACGACCGCTTCGCTCGATGAGCAACCTCAAT	900
QY	901	CTGCAAGACTCGACATGCTCTCGTCTCTTCTCGGAGGCTTGAACGCTCTCAAGGTGAG	960
Db	901	CTGCAAGACTCGACATGCTCTCGTCTCTTCTCGGAGGCTTGAACGCTCTCAAGGTGAG	960
QY	961	GCCATGCTCGGACAGAGGCTCTTTCGCGCCGTTTCACTACAGACTCTGCGCCCGAC	1020
Db	961	GCCATGCTCGGACAGAGGCTCTTTCGCGCCGTTTCACTACAGACTCTGCGCCCGAC	1020
QY	1021	CCCGGCAAGTGAAGTGGCGGCAACTCCGACGCTCTTTCGCGCTGTGCTTTGCC	1080
Db	1021	CCCGGCAAGTGAAGTGGCGGCAACTCCGACGCTCTTTCGCGCTGTGCTTTGCC	1080

QY	108	GTTCAGACAGAGGAGGATCAAGGTCAAAGGACACAGAGGCAATTCCTCGCCAGACCCG	114
Db	1081	GTTTAGACAGAGAGAGGATCAAGGTCAAAGACAGAGGCAATTCCTCGCCAGACCCG	114
QY	1141	TACCCGCTCCGACGTCCTCAGTTCTCTGACCCTGCTGTGAGGACATGATGACCGC	120
Db	1141	TACCCGCTCCGACGTCCTCAGTTCTCTGACCCTGCTGTGAGGACATGATGACCGC	120
QY	1201	TACTCGACTCTCTCGCTCGAAGAACACACACACACACACACACACACACACACACAC	126
Db	1201	TACTCGACTCTCTCGCTCGAAGAACACACACACACACACACACACACACACACACAC	126
QY	1261	AACAAGACAGCCGCGACGCGCGGCAACTTCCAGGCGCTGCGCTGCTGATTTGATGAGAG	132
Db	1261	AACAAGACAGCCGCGACGCGCGGCAACTTCCAGGCGCTGCGCTGCTGATTTGATGAGAG	132
QY	1321	AAGACCAAGGCTCGCACTCGCCTCATTCGGCAAGCTCAATTCAACGATGACACCAATTG	138
Db	1321	AAGACCAAGGCTCGCACTCGCCTCATTCGGCAAGCTCAATTCAACGATGACACCAATTG	138
QY	1381	CTGAACGCTGACATGAACCGGCGCCTGCTGTGCTCGCTGACGAGGACCCGTCGCTC	144
Db	1381	CTGAACGCTGACATGAACCGGCGCCTGCTGTGCTCGCTGACGAGGACCCGTCGCTC	144
QY	1441	AACATACACGAGGAGGCTTGAGCAATTCACATCGCTGCTTACGCTTCGAGCTCGGACAC	150
Db	1441	AACATACACGAGGAGGCTTGAGCAATTCACATCGCTGCTTACGCTTCGAGCTCGGACAC	150
QY	1501	CTTGCCAAACCCGCTGACATCACTTTTGCTCCAGCCGACAGATGAGGCAACCAAGCCGTCAAC	156
Db	1501	CTTGCCAAACCCGCTGACATCACTTTTGCTCCAGCCGACAGATGAGGCAACCAAGCCGTCAAC	156
QY	1561	TGCGCTGCTCTATCTCCGCGCGCGGCACTGCGAGAGCAACACAGCTCTTCTCTCTCT	162
Db	1561	TGCGCTGCTCTATCTCCGCGCGCGGCACTGCGAGAGCAACACAGCTCTTCTCTCTCT	162
QY	1621	CTCGCTCGCACCTGTATCTGACGCTCAAGCCGTGACCTTCGCGCGATGAGAGCTCGAC	168
Db	1621	CTCGCTCGCACCTGTATCTGACGCTCAAGCCGTGACCTTCGCGCGATGAGAGCTCGAC	168
QY	1681	TTCAAGAGACAGTTGACACCGCTTCTCCGACCTCTCTCAAGAGACCTCGGCACTGGC	174
Db	1681	TTCAAGAGACAGTTGACACCGCTTCTCCGACCTCTCTCAAGAGACCTCGGCACTGGC	174
QY	1741	CTCGACGTCAACGCACTTGCGCTTGAGGTCAGAGGCGCTCAACAGCGTCTCGACAG	180
Db	1741	CTCGACGTCAACGCACTTGCGCTTGAGGTCAGAGGCGCTCAACAGCGTCTCGACAG	180
QY	1801	ACGACGAGGTACGACCTCGACGCGCGCTGCGACAGAGCCTTCTGTAAGCGACCGGACCC	186
Db	1801	ACGACGAGGTACGACCTCGACGCGCGCTGCGACAGAGCCTTCTGTAAGCGACCGGACCC	186
QY	1861	GTCGTGAGCTCTCTGCTCGCTCGCCCTTGCCCAAGTCACCTTACTGCGGTCAACGCG	192
Db	1861	GTCGTGAGCTCTCTGCTCGCTCGCCCTTGCCCAAGTCACCTTACTGCGGTCAACGCG	192
QY	1921	TGGAAGGTTGCTCGCGCGCGAGAAAGCCATTCGCTTACGCGGAGGTGCGAACCGCTTC	198
Db	1921	TGGAAGGTTGCTCGCGCGCGAGAAAGCCATTCGCTTACGCGGAGGTGCGAACCGCTTC	198
QY	1981	TGGCAGACGCGCTTTCGACAGGCGCGGCGACAGCACTCTTCGCGCGACCGCGCTC	204
Db	1981	TGGCAGACGCGCTTTCGACAGGCGCGGCGACAGCACTCTTCGCGCGACCGCGCTC	204
QY	2041	CTGTACTCTGTTCTGTCGCGAGAGCTCGGCGTGCAGAGCGCGCGCGACGCTGTTGTC	210
Db	2041	CTGTACTCTGTTCTGTCGCGAGAGCTCGGCGTGCAGAGCGCGCGCGACGCTGTTGTC	210
QY	2101	GCGCGGACAGAGAGAGATCGGAGACCAAGCTCTGCGGCAATTCAGAGCCATCAAGAC	216
Db	2101	GCGCGGACAGAGAGAGATCGGAGACCAAGCTCTGCGGCAATTCAGAGCCATCAAGAC	216
QY	2161	GCGCGCAACCAACGCTCTGTCAGAGATCTCGGTAAAGCCGAGCAACCTCGCTA	222

Db	2161	GGCCGCATCAACCAAGTCTCTGGTCAAAATGCTCGCGTAAGGCCCAAGCAAGCTCGCTCA	2220
Qy	2221	GACGCCCGCTCACCCTCAAGACACAGCTTTTCGACGTGTGTGTGTGGCAAGAACGACTTT	2280
Db	2221	GACGCCCGCTCACCCTCAAGACACAGCTTTTCGACGTGTGTGTGTGGCAAGAACGACTTT	2280
Qy	2281	CCTCCATACACATGTGTGCGCTTACTCTCTCGCGCGTCATCAGTCTCTCAAGTTCTTTCGAT	2340
Db	2281	CCTCCATACACATGTGTGCGCTTACTCTCTCGCGCGTCATCAGTCTCTCAAGTTCTTTCGAT	2340
Qy	2341	CCGCGCCTCTCGGT	2400
Db	2341	CCGCGCCTCTCGGT	2400
Qy	2401	AGTTCAAAAAAAAAAAAAA	2419
Db	2401	AGTTCAAAAAAAAAAAAAA	2419

CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 29
LENGTH: 2163
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Preferred
OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
NAME/KEY: CDS
LOCATION: (1)..(2163)
OTHER INFORMATION:
US-09-939-408A-29

Query Match 70.4%; Score 1704; DB 9; Length 2163;
Best Local Similarity 84.0%; Fred. No. 0;
Matches 1817; Conservative 109; Mismatches 236; Indels 0; Gaps 0;

37 ATGGCCCTTCCTTGAGTCTGCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 96
1 ATGGCCCTTCCTTGAGTCTGCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 60
97 CACGCGCTCGCCACAGTCTGCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 156
61 CACGCGCTCGCCACAGTCTGCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 120
157 CTCGATGCGACGCGCGCCACAGTCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 216
121 CTCGATGCGACGCGCGCCACAGTCTGCGCCACAGCTCGCCACAGGCTTTACCAACGAGCTCG 180
217 CCCACGACGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 276
181 CCCACGACGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 240
277 GCGCGCGCCACAGGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 336
241 GCGCGCGCCACAGGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 300
337 GACGAGGCTGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 396
301 GACGAGGCTGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 360
397 GGTTCGCTGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 456
361 GGTTCGCTGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 420
457 ATCGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 516
421 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 480
517 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 576
481 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 540
577 TCGCTACGCGCGCCACAGTCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 636
541 TCGCTACGCGCGCCACAGTCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 600
637 TTGAACGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 656
601 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 660
697 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 756
661 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 720
757 GTTTTCGACGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 816

721 GTTTCGACGAGGACGCTGCTGCTGCGCCACAGTCTGCGCCACAGGCTTTACCAACGAGCTCG 780
817 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 876
761 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 840
877 TCGGCTGATGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 936
841 TCGGCTGATGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 900
937 GCTTTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 996
901 GCGCTACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 960
997 ATTCAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1056
961 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1020
1057 CTCCTTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1116
1021 CTCCTTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1080
1117 GAGGACATTTCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1176
1081 GAGGACATTTCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1140
1177 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1236
1141 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1200
1237 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1296
1201 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1260
1287 TCGGCTGCTGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1356
1261 DCGGCTGCTGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1320
1357 AACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1416
1321 AACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1380
1417 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1476
1381 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1440
1477 GCTTTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1536
1441 GCTTTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1500
1537 GAGATGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1596
1501 GAGATGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1560
1597 GCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1656
1561 GCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1620
1657 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1716
1621 GACGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1680
1717 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1776
1681 CTCGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1740
1777 GCGGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1836
1741 DCGGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1800
1837 GCGGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1896
1801 GCGGAGGACGCTGCTGCTGCGCCACAGTCTGCTGCGCCACAGTCTGCTGCGCCACAGGCTTTACCAACGAGCTCG 1860

QY 1897 GTACACCTTACTGCGCTCAACGCGTGAAGAGTTGGCTCGCGGAGAAAGCCATCTGCTC 1956
 DB 1861 GTCTCGCTGCBCCGCTTCAACGCTGAAGAGTGGCTCGCGGAGAAAGCCATCTGCTC 1920
 QY 1957 ACCGCGAGAGTGGCAACCGCTTCTGGCAGACCGCTCTTGGCAGCGCCCGCGCAGCA 2016
 DB 1921 ACCGCGBAGTGGCAACCGCTTCTGGCAGACCGCTCTTGGCAGCGCCCGCGCAGCA 1980
 QY 2017 TACCTCTCGCGCGCGACGCGCTGTACTGTCTGTTGCGGAGAGAGAGTGGAGTGGAG 2076
 DB 1981 TACCTCTCGCGCGCGACGCGCTGTACTGTCTGTTGCGGAGAGAGTGGAGTGGAG 2040
 QY 2077 GCGCGCGCGCGCGACGCTGTTTGTGCGCGTGAAGAGAGAGAGTGGAGAGAGTGGAG 2136
 DB 2041 GCGCGCGCGCGCGACGCTGTTTGTGCGCGTGAAGAGAGAGAGTGGAGAGAGTGGAG 2100
 QY 2137 CGCATCTACGAGGCGCATCAAGAGAGAGAGAGAGAGTGGAGAGAGTGGAG 2196
 DB 2101 CGCATCTACGAGGCGCATCAAGAGAGAGAGAGAGAGTGGAGAGAGTGGAG 2160
 QY 2197 TA 2198
 DB 2161 TA 2162

RESULT 4
 US-09-939-408A-20
 Sequence 20, Application US/09939408A
 Patent No. US20020102712A1
 GENERAL INFORMATION:
 APPLICANT: Yoshida, Robera
 TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 FILE REFERENCE: 29479/50NSCA
 CURRENT APPLICATION NUMBER: US/09/939,408A
 CURRENT FILING DATE: 2001-08-24
 PRIOR APPLICATION NUMBER: US 09/624,693
 PRIOR FILING DATE: 2000-07-24
 PRIOR APPLICATION NUMBER: PCT/US01/23270
 PRIOR FILING DATE: 2001-07-24
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 20
 LENGTH: 2475
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 NAME/KEY: misc difference
 LOCATION: (13)-(2475)
 OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
 OTHER INFORMATION: Description of Artificial Sequence: Consensus
 OTHER INFORMATION: Sequence of SEQ ID NOS: 12, 16, and 18
 US-09-939-408A-20

Query Match 67.3% Score 1629; DB 9; Length 2475;
 Best Local Similarity 78.0%; Pred. No. 0; Mismatches 504; Indels 25; Gaps 8;
 Matches 1878; Conservative 1;
 QY 37 ATGGCCCTTCTGAGTCTGCTGCGCAACGAGTCTGCGCAACGAGCTTACCAACGAGCTG 96
 DB 1 ATGGCCCTTCTGAGTCTGCTGCGCAACGAGTCTGCGCAACGAGCTTACCAACGAGCTG 60
 QY 97 CACGCGGCTCGCAACGAGTCTGCGCGCGCGCTTCTGCGCTCTGCGCGCGCAAGC--CG 154
 DB 61 CACGCGGCTCGCAACGAGTCTGCGCGCGCGCTTCTGCGCTCTGCGCGCGCAAGC--CG 120
 QY 155 GCGTGAATGAGCAACCGCGCGCAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 214
 DB 121 GCGTGAATGAGCAACCGCGCGCAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 180

QY 215 ACCCGACGAC---GACGTCTGAGCTCAGCGGGATACAGCTTCAACCGCTGAGCTTG 271
 DB 181 ACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240
 QY 272 TCG 330
 DB 241 TCG 300
 QY 331 CGCGTGAACGAGGCTGAGCTTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
 DB 301 AAGATGCAACGAGGCTGAGCTTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 QY 391 ACCAGGAGTTCGAGTCTGCG 450
 DB 361 ACCAGGAGTTCGAGTCTGCG 420
 QY 451 GCGTGAATGAGCAACCG 510
 DB 421 GCGTGAATGAGCAACCG 480
 QY 511 GCGTGAATGAGCAACCG 570
 DB 481 GCGTGAATGAGCAACCG 540
 QY 571 GTCAACTGCTCAGCG 630
 DB 541 GTCAACTGCTCAGCG 600
 QY 631 AACTCTTGAACGACG 690
 DB 601 AACTCTTGAACGACG 660
 QY 691 GCGTGAATGAGCAACCG 750
 DB 661 GCGTGAATGAGCAACCG 720
 QY 751 GTCAACTGCTCAGCG 810
 DB 721 GTCAACTGCTCAGCG 780
 QY 811 TTTGAGTCTGAGGCG 870
 DB 781 TTTGAGTCTGAGGCG 840
 QY 871 GCGTGAATGAGCAACCG 930
 DB 841 GCGTGAATGAGCAACCG 900
 QY 931 TCGAGGCG 990
 DB 901 TCGAGGCG 960
 QY 991 CGGTTCAATGAGCAACCG 1050
 DB 961 CGGTTCAATGAGCAACCG 1020
 QY 1051 CGGTTCAATGAGCAACCG 1110
 DB 1021 CGGTTCAATGAGCAACCG 1080
 QY 1111 GAGCAGAGGAGCAATCTTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1170
 DB 1081 GAGCAGAGGAGCAATCTTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
 QY 1171 GCGTGAATGAGCAACCG 1227
 DB 1141 GCGTGAATGAGCAACCG 1200
 QY 1228 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1287
 DB 1201 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1260
 QY 1288 TTCCAGGCGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1347

1261 TTCCAGGCGGCGCTGTCGCGAACAACATGAGAGAACTGCGCTCCGCTGCGCCCTGATC 1320
1348 GGCAAGCTCAACTTCAGCAGAGTGCACCGAGTGTGTCACAGCTCCATGAACCGCGGCTG 1407
1321 GGCAAGCTCAACTTCAGCAGAGTGCACCGAGTGTGTCACAGCTCCATGAACCGCGGCTG 1380
1408 CTTTGTGCTGCTGCTGCGGAGGAGCCGCTGCTCAACTATCAAGGCGCAAGGCTTGAATT 1467
1381 CCGTCTGCTGCTGCTGCGGAGGAGCCGCTGCTCAACTATCAAGGCGCTGCAATT 1440
1468 CACATGCTGCTTACGCTTGGAGCTGCGGACCTGTCGCAACCGGCTCACTACTCTGCTG 1527
1441 GCGCGCGCGCTTCACTTGGAGCTGCGGACCTTCTGCAACCGGCTGCACTGCTGCTGCT 1500
1528 CAGCGCGGAGAGTGGGAGCAACAGCGGCTCAACTGCTGCTGCTCACTTCTGCGCGCGG 1587
1501 CAGCGCGGAGAGTGGGAGCAACAGCGGCTCAACTGCTGCTGCTCACTTCTGCGCGCGG 1560
1588 ACTGCGGAGGCGCAACAGAGCTCTTCTGCTGCTTCTGCTGCTGCACTGCTGCAAGCTG 1647
1561 ACNCGCGAGGCGCAACAGAGCTCTTCTGCTGCTGCTGCTGCACTGCTGCTGCTGCTG 1620
1648 CAGCGCGGAGAGTGGGAGCAACAGCGGCTCAACTGCTGCTGCTCACTTCTGCGCGCTG 1707
1621 CAGCGCGGAGAGTGGGAGCAACAGCGGCTCAACTGCTGCTGCTCACTTCTGCGCGCTG 1680
1708 CCGACTCTCTCCAGCAGACCTCGGCACTGCGCTGAGCTC-----AACGA 1755
1681 CCGACTCTCTCCAGCAGACCTCGGCACTGCGCTGAGCTC-----AACGA 1740
1756 CTTGCGCTCGAGTCAAGAGGCGCTCAACAGAGTGTGTCAGAGAGAGAGAGAGTGCAG 1815
1741 CTGAGGAGCAAGGCTCAAGAGGCGCTCAACAGAGGCTCGAGAGAGAGAGAGTGCAG 1800
1846 CTGAGGCGGCGCTGCGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1875
1801 CTGAGGCGGCGCTGCGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
1876 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1933
1861 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1934 CCGCGGAGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1993
1921 CCGCGGAGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1981 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
1994 CTTGCGAGGCGCGCGGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2053
2054 TCGCGAGAGAGGCTGCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2113
2041 TCGCGAGAGAGGCTGCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
2114 AACAGATGGAGAGAGGCTGCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2173
2101 TGAAGATGGAGAGAGGCTGCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2160
2174 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2232
2161 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
2233 ACCCGAAGAGAGGCTTTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2292
2221 CCGCGAGAGAGGCTTTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
2293 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2352
2281 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
2353 GGTGCTCA-GTACAGGTGTATAGAGCTGGAATGATTTGCAAGCTTTCAGAGTTCAAAAA 2411

Db 2341 NGTNNNCAANNACCTNN 2400
Qy 2412 AAAAAAAAA 2419
Db 2401 NNNNNAAA 2408
RESULT 5
US-09-765-873A-9
Sequence 9, Application US/09765873A
Patent No. US20010053847A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: ECI009 US CIP
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO: 9
LENGTH: 2151
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2151)
US-09-765-873A-9
Query Match 49.5%; Score 1196.6; DB 9; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2,4e-282;
Matches 165; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
Qy 99 CGCGCTTCGACCAAGTGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 158
Db 45 CGCGCATCGCAAGAGAGGCTGCAATGAGCGCTGCAACCACTGCGAGTCCAGAGCTTC 104
Qy 159 CGATGGCCAGCGCGCGGAGCAAGTGGAGCTCGAGTGGTCAAGAGTCTTCAGAGAGCC 218
Db 105 GCACCTGCGCAACCAAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 164
Qy 219 CACCGAGAGCTGCTGCGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 278
Db 165 GACGATCGAGAGCTGCGAGCTGCGAGCTGCGAGCTGCGAGCTGCGAGCTGCGAGCTG 224
Qy 279 CGCGCGAAGGAGGCGAGAGGCTGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
Db 225 CGCGAGAGAGGAGGAGGCTGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
Qy 339 CAGAGAGCTGCACTTCTCAAGGCGCGGCTTCAAGACTCGGCTGAGAGAGTCAACAGAG 398
Db 285 CAATGAGTGGAGTCTTTCGCGCTGCGAGACTCTCAATAGAGTCTCAAGAGTCTCAAGAG 344
Qy 399 TTTGAGTGGCTGCGGCGAGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 458
Db 345 ATTGGCGAGATCGGAG 404
Qy 459 CGAGAGAGAGCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Db 405 CGAGAGAGAGCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
Qy 519 CCGTGAAGAGAGAGCTTTCGCGCTGAGAGTGGTGGCGGCGGAGAGAGAGAGAGAGAGAG 578
Db 465 TCTCGAGAGAGTGGCTTTCGCGCTGAGAGTGGTGGCGGCGGAGAGAGAGAGAGAGAGAG 524
Qy 579 GCTCAAGGAG 638
Db 525 CTTGAGCGGCGGAG 584
Qy 639 GAACACCGAG 698

Db 585 CAACCAAGGATACCCCATGTCCTCCGCGGACCACTCTGTGGGGGAGCT 644
 Qy 699 CAGCCGCTCTCTAGATGCGCGGCGCATCAACCGGTACCCCGCATCAAGTTCAAGT 758
 Db 645 CTCTCTCTCTCTATGTCAGCGGCTTACGGGTACCCGGACAGCAAGGTGACGT 704
 Qy 759 TTTCAGAGGGAACCGAAGATCATGTTTGGCGGAGAGCCATCTGCTTTGGTCT 818
 Db 705 CGTCCAGAGGGAAGAGATCTGTACCGCCGAGGAGATGCGCTTTCAACT 764
 Qy 819 CGAGGAGTGTCTTGGGCGGAGAGGAGTCTGGTGTGTCACAGGAAAGGCGGTCTC 878
 Db 765 CGAGCCGCTGTCTCTGCGCCCAAGAGAGGTCTGGGTCTGTCAACGAGCAACGCGTCTC 824
 Qy 879 CGCTGTATGGCGACCTTCAATGTCTGACAGCTGCGACATGCTCTGCTCTCTGCAAGC 938
 Db 825 AGCATGATGGCCACCTCGCTCTGCAAGAGCAACATGCTCTGCTCTCTGCAAGT 884
 Qy 939 CTTCAGGCTCTCAAGTGGAGGAGCATGATGGGCTGAGAGGCTTCTGCGCGCTCAT 998
 Db 885 GCTCAAGGCAATGACGCTTGAAGCGATGAGCGACCGCGCTGTTCACACCTCTCTCT 944
 Qy 999 CCAAGAGTCTGCGCGCGGACCCGCGCAGGTGAGGTGCGCGGCAACATCCGACGCT 1058
 Db 945 TCACGAGCTCAAGCGCCCTCACCCGACGAGATGMAATCGGGGAAACATCCGACGCT 1004
 Qy 1059 CCTTTCGGCTGTGCTGTTTCCGTTGAGCAAGAGAGAGGTCAAGTCAAGAGAGCA 1118
 Db 1005 CCTCGAGGAGACCGCTTGTGTGTCTCACATGAGAGAGGTCAAGTCAAGAGAGCA 1064
 Qy 1119 GGGCATTTCTTGGCGAGACCGCTACCGGCTCCGACGCTGAGTTCTTGGGCGGCT 1178
 Db 1065 GGGCATTTCTTGGCGAGACCGCTACCGGCTTGGGAGCGTCTTCTGAGTGGCGCGCT 1124
 Qy 1179 CGTGAAGACATGATGACAGCGCTACTCTGCTCTGCTGAGAAACA---GACGAC 1235
 Db 1125 CGTGAAGACCTCATTTCAAGGCCACGCGCTCTCAACATGAGGCGCGGACATGAGAC 1184
 Qy 1236 CGAACAACCGCTCTCTGAGCTGAGAAACAGCAACCGCGGAGCGGCAATTCAGAGC 1295
 Db 1185 CGAACAACCTCTCATGAGCTGAGAAACAGCAATTCACACCGCGGCAATTCAGAGC 1244
 Qy 1296 GTGCGCTGTCTGATTTGATGAGAAAGACAGGCTGCACTGCGCTCTCAAGAGCT 1355
 Db 1245 TGGCGCTGTGGCCACACATGAGAAAGCTCGCTGCGCTCGCCAGATCGCAAGCT 1304
 Qy 1356 CAATTTACAGAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCTGCTTCTG 1415
 Db 1305 CAATTTACAGAGTGCACCGAGATGCTCAACGCGGATGAACCGCGGCTGCTTCTG 1364
 Qy 1416 CTTGCTGCGAGGAGCGGCTGCTCAACTATGAGGAGAGGCTTGAACATTCACATGCT 1475
 Db 1365 CTTGCGGCGGAGAGCGGCTGCTCTCTTACACTGAAAGGCTTGCACATGCGCGCTCTC 1424
 Qy 1476 TGCTTACGCTTTCGAGCTGCGCACTTGGCAACCGGCTCACTACTCTGTCAGAGCGC 1535
 Db 1425 GGGCTACACCTCGAGGTTGGAGACCTGCGCAACCTGTACAGAGAGTTCAGCGGCG 1484
 Qy 1536 AGAGATGGGCAACAGGCGCTCACTGCTGCTCTATCTTCCGCGGCGCGCATCTGCGA 1595
 Db 1485 TGAAGTGGGCAACAGGCGGCTCACTGCTGCTCTATCTGCGCTCTGTCGACGACGA 1544
 Qy 1596 GGGCAACAGCTCTCTTCTCTCTCTCTGCTGCTGCACTGACGCTCAGAGCGCT 1655
 Db 1545 GTTCAAGAGAGCTCTTCTCTCTCTCTCTGCGCAACCTTACTGCTTCTCAAGAGCT 1604
 Qy 1656 CGACTCTGCGGAGTGAAGCTGCACTTCAAGAAAGCAAGTGAACCGCTTCTTCCAGCTT 1715
 Db 1605 CGACTTGGCGGAGCGAGTTCAGTTTCAAGAAAGCAAGTGAACCGCTTCTTCCAGCT 1664
 Qy 1716 CTTCAAGAGAGCTTGGCACTGAGCTGCAAGTCA-----ACGCACTTGGCGCT 1763

Db 1665 CATGACAGCACTTTGGCTCCGCCATGACGGGCTCGAAGCTTGGCGAGAGCTGTGCA 1724
 Qy 1764 CGAGTCAAGAGAGGCGCTCAACAGGCTCTGAGACAGACAGACGTAAGACCTCGAGCC 1823
 Db 1725 GAAAGTGAACAGAGCGCTCGCCAAAGCGCTTGACACACCACTGTACAGCTCTCTCC 1784
 Qy 1824 GCGCTGGACAGACGCTTCTTGTGAAGAGCGAGCGGAGCGGTGTGAGCTCTCTCTCTC 1883
 Db 1785 GCGCTGGACAGCGCTTC 1844
 Qy 1884 GCGCT 1943
 Db 1845 GTC-----GCTCTGCTGCGCGCTGTCACAGCTTGAAGGTGCGCGCGCGAGTCTC 1895
 Qy 1944 GGCATCTCTGCTCAAGCGCGAGAGTGCACACCGCTTCTGCAACGCGCTTCTGCAAGC 2003
 Db 1896 GGCATCTCTGCTCAAGCGCGAGAGTGCACACCGCTTCTGCAACGCGCTTCTGCAAGC 1955
 Qy 2004 GCGGCGGACGATCT 2063
 Db 1956 GCGGCGGCTCTGATCT 2015
 Qy 2064 GCTGAGGCTGAGGCGCGCGCGCGGCGGAGCTTGTGTGCGGTGTCAGACAGAGATGCG 2123
 Db 2016 GCTTGGCTCAAGCGCGCGCGCGGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2075
 Qy 2124 GAGCAAGCTCTGCGGATCTTACGAGGCTCAAGAGCGCGGCGATCAACAGCTCTCTCT 2183
 Db 2076 CTCGAAAGCTTCCAAATCTTACGAGGCTCAAGGCTCAAGGAGATCAACAGCTCTCTCT 2135
 Qy 2184 CAAGATGTCGCGTA 2198
 Db 2136 CAAGATGTCGCTTA 2150

RESULT 6
 US-10-138-970A-23
 ; Sequence 23, Application US/10138970A
 ; Publication No. US2003079255A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sarsilani, Sima
 ; APPLICANT: Tang, Xiao-Song
 ; APPLICANT: Qi, Wei Wei
 ; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
 ; TITLE OF INVENTION: hydroxycinnamic Acid
 ; FILE REFERENCE: CUI777
 ; CURRENT APPLICATION NUMBER: US/10/138,970A
 ; CURRENT FILING DATE: 2002-07-23
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 23
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: mutant from Rhodotorula glutinis
 US-10-138-970A-23

Query Match 49.5%; Score 1196.6; DB 15; Length 2151;
 Best Local Similarity 74.0%; Pred. No. 2,4e-282;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

Qy 99 CGCGGCTCGACCAAGTGGGCTGGGAGCGGCACTTGGGCTTGGGCGGAGCGCGGCT 158
 Db 45 CTGCGATCTCGCAAGAGGCTGTCAATGCGGCTTGCACCACTTGCAGTGGAGGCTC 104
 Qy 159 CGATGGCAGCGCGGCGACAGTGCAGCTGAGATGCTGAGAGAGCTTCTTAAGGAGCC 218
 Db 105 GCACCTGCGCCACCAACCAAGGTGACAGGATGACATCTGAGAGAGATGCTTGGCGGCGC 164
 Qy 219 CACGAGCAAGTGTGAGAGTCAAGGAGTCAAGGCTCAAGGCTTGGAGGCTTGGAGGCT 278
 Db 165 GACGAGTCAAGGCTGAGAGCTGAGGCTTCTGCTCAACTTGGAGAGGCTGCTGGC 224

PRIOR APPLICATION NUMBER: US 60/147,719
 PRIOR FILING DATE: 1998-08-06
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 9
 LENGTH: 2151
 TYPE: DNA
 ORGANISM: Rhodotorula glutinis mutant
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(2151)
 OTHER INFORMATION:
 US-10-188-523B-9

Query Match 49.5% Score 1196.6; DB 15; Length 2151;
 Best Local Similarity 74.0%; Pred. No. 2.4e-282;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 GCGCGCTCGACCAAGTCGGCTGCGGCGCCCACTTCGCTTCGCGCGACGCGCGCT 158
 DB 45 CCGCGATCCGCAAGCAGGCTGTCAATGCGCCCTCGACCACTCGCAGTCGAGGCTC 104
 QY 159 CGATGGCCACGCGCGCCGACCACTCGCAGCTGCAATGTCGAGAGCTCTTCAAGCC 218
 DB 105 GCACTGCGCCCAACCCCAAGTCAACGAGTCACTGTGAGAAAGTGTCTCCCGGCT 164
 QY 219 CACCGACGACGTCTGAGCTCAGCGGGTCAAGCCTCAAGCTTCGAGCTTGTGAGCG 278
 DB 165 GACCGACTGACGCTCGAAGTCAAGGCTTCTGCTCACTCGGACCTCGGACCTCTCGG 224
 QY 279 GCGCGCAAGGCGCGAGGTCGCGCTCAGAGCAGCAGAGATCCGCGACGCTCGA 338
 DB 225 CCGCGAAAGGCGCGGCTGTCTCGCTCAAGACAGCAGAGATCCGCTCAAGATTGA 284
 QY 339 CAAGAGCGTCGACTTCTCAAGGCGCGCTTCAAGATCCGCTCAAGGATCCAGCGG 398
 DB 285 CAATCGCTGAGTCTTGGCGCTCGCACTCTCCATAGAGCTTACGGGCTCAAGCTGG 344
 QY 399 TTTCGGTGTCTGCGCGACACAGAGACTGAGATGAGTCAAGCTTCGAAAGGCGCTAT 458
 DB 345 ATTTGGCGGATCCGAGACACCCGACCGAGAGAGCCATCTCGCTCAAGAGGCTCTCT 404
 QY 459 CGAGACGACGCTTCTCGCGCGTGAAGCGGCTCGGCTCTCGCTTCAAGGCTCGGAG 518
 DB 405 CGAGACGACGCTTCTCGCGCGTGAAGCGGCTCGGCTCTCGCTTCAAGGCTCGGAG 464
 QY 519 CCTCGAGAACAGCTTCTCGCTGAGGTCGTCGCGCGCGCATGATCCTCGGCTCAATC 578
 DB 465 TCTCGAGAACTGCTTCTCGCTGAGGTCGTCGCGCGCGCATGATCCTCGGCTCAAG 524
 QY 579 GGTCAAGGTCGCTTCTCGCGCGTCTCGCTGTCGCTTGAAGGCTTCAAGCTTCTT 638
 DB 525 CTTCGCGCGCGCATGCTGCTGTCGCTGTCGCTGTCGCTGTCGAGCGCTCAACACTTCT 584
 QY 639 GAACACCGCATCAAGCCCATCTGCTTCTCGCGCGCTCATCTCGCGCTCGGAGCT 698
 DB 585 CAACACCGCATCAAGCCCATCTGCTTCTCGCGCGCTCATCTGCTGCTGCTGCTGCT 644
 QY 699 CAGCGCTCTGTCATCTGCGCGCGCATCAAGGTCACCCGACGTCAGAGTCAAGTTCAGT 758
 DB 645 CTCTCTCTCTCTCAATGTCAGCGCGCATCAAGGTCACCCGACAGAGGTCAGCT 704
 QY 759 TTTCGACGAGGAAACGAGATCATGTTTTCGCGCGAGGCGCATCTGCTTCTTGGTCT 818
 DB 705 GGTTCACGAGGAGGAGAGAGATCTCTGACCCCGAGGCGCATGAGGCTCTTCAACT 764
 QY 819 CGAGCACTGCTCTCTGCGCGCGAGAGGCTTCTGCTTCTGTCAGAGAGAGCGGCTCTC 878
 DB 765 CGAGCGCTGCTCTCTGCGCGCGAGAGGCTTCTGCTTCTGTCAGAGAGAGCGGCTCTC 824
 QY 879 GCGCTCATGCGCGCGCTGTCAGAGCTGTCAGAGCTGTCAGAGCTGTCAGAGCTGTCAGAG 938
 DB 825 AGCATGATGCGCGCTGCTGTCAGAGCTGTCAGAGCTGTCAGAGCTGTCAGAGCTGTCAGAG 884

QY 939 CTTGACGCTCTACGCTGAGGCGCATGCTGCGCGACAGAGGCTGTCGCGCTTCAAT 998
 DB 885 GCTACGCGCATGACGCTGAGAGGATGTCGCGCGACGCGGCTGTCACCCCTTCTCT 944
 QY 999 CCAGACGCTGCGCGCGCGACCCCGGCGAGTCTGAGGTCGCGCGCAATCTCGACGCT 1058
 DB 945 TCAGAGCTACGCGCGCTTACCCGAGCAGATCGAAGTCGCGGAAACATCTCGAGCT 1004
 QY 1059 CTTTCGCGCTGCTGCTTTCGCTTTCGAGCAGAGAGAGTCAAGTCAAGAGCAGCA 1118
 DB 1005 CTTGAGAGAGAGCGCTTCTCTCTGTCACCATGAGAGAGGTCAGAGTCAAGAGCAGCA 1064
 QY 1119 GGGCATTTCTGCGAGAGAGCGTACCGGCTCGGCGAGCTGCTGCTGCTGCTGCTGCT 1178
 DB 1065 GGGCATTTCTGCGAGAGAGCGTACCGGCTCGGCGAGCTTCTCTGAGTCTGCTGCTGCT 1124
 QY 1179 GGTGAGAGATGATGACAGCGCTTCTGACTCTGCTGCTGAGAACACA---CGAGCAG 1235
 DB 1125 GGTGAGAGATGATGACAGCGCTTCTGACTCTGCTGCTGAGAACACA---CGAGCAG 1184
 QY 1236 CGACAAACCGCTCTGAGCTGAGAGAACAGAGACCGGCGAGGCGGCACTTTCAGAGC 1295
 DB 1185 CGACAAACCGCTCTGAGCTGAGAGAACAGAGACTTTCGACACAGGCGGCAATTTTCAGAGC 1244
 QY 1296 GTGCGCTGCTGCTTCTGATGAGAGAACAGAGGCTCGGACTCGGCTCATGCGAGAGCT 1355
 DB 1245 TGCGGCTGTGCGCAACACCATGAGAGAGAGTCTGCGGCTCGGCGAGTCTGCGAGAGT 1304
 QY 1356 CAATTTACGAGATGACACGAGTTCCTCAAGCTGTCATGAAACCGGCGCTTCTGCTG 1415
 DB 1305 CAATTTACGAGATGACACGAGTTCCTCAAGCTGTCATGAAACCGGCGCTTCTGCTG 1364
 QY 1416 CTTGCGCGAGAGACCGGCTGCTGCTCAATGACAGGAGGCTTGAATTCATGATGAGC 1475
 DB 1365 CTTGCGCGAGAGACCGGCTGCTGCTCAATGACAGGAGGCTTGAATTCATGATGAGC 1424
 QY 1476 TGCTTACGCTTCTGAGGCTGCGGCACTTTCGCAACCGGCTCAATCTTCTGTCAGAGC 1535
 DB 1425 GAGCTTACCTCTGAGGTCGAGACCTTCTGCAACCGGCTCAATCTTCTGTCAGAGC 1484
 QY 1536 AGAGATGAGGAAACAGAGCGGTCATCTGCTGCTCAATCTTCTGTCAGAGC 1595
 DB 1485 TGAATGAGGAAACAGAGCGGTCATCTGCTGCTCAATCTTCTGTCAGAGC 1544
 QY 1596 GGGCAAGAGCTCTTCTCTCTTCTGCTTCTGCTGCACTTTCGCAACCGGCTCAAGGCT 1655
 DB 1545 GTCGAAGAGCTCTTCTCTCTTCTGCTTCTGCTGCAACCGGCTCAAGGCT 1604
 QY 1656 GAGCTTCTGCGGAGTGAAGTCTGACTTCAAGAGAGTTCGACCGGCTTCTGCGAGCT 1715
 DB 1605 GAGCTTCTGCGGAGTGAAGTCTGACTTCAAGAGAGTTCGACCGGCTTCTGCGAGCT 1664
 QY 1716 CTTGAGAGAGCTTCTGAGAGTCTGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1763
 DB 1665 CATGAGAGAGCTTCTGAGAGTCTGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1724
 QY 1764 CGAGGTCAGAGAGGCTTCAAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
 DB 1725 GAAAGTGAAG 1784
 QY 1824 GCGCTGAG 1883
 DB 1785 GCGCTGAG 1844
 QY 1884 GCGCTTCTGCAAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
 DB 1845 GTC-----GCTCTGCTGCGCGCGCTCAAGCTTGAAGAGAGAGAGAGAGAGAG 1895
 QY 1944 GCGCATCTGCTGAG 2003
 DB 1896 GCGCATCTGCTGAG 1955

2004 GCCGGCGGACGATCTCTTCGCGCGGACAGCGCGCTCTGTACTGCTTGTGGCGGAGGA 2063
2063
1956 GCCGCGCCTCTCGAATCTTCGCGCGGCACTGAGATCTCTAGCCTTGTGTCGCGGAGGA 2015
2015
2064 GCTCGGCGTGAAGGCGCGCGCGGCGGACGCTGTTGTGCGGCTGACAGAGACGATCGG 2123
2123
2016 GCTTGGGCTAAGGCGCGCGCGGAGAGCTTCTCTCGGAGGAGAGGTAGCATGG 2075
2075
2124 GAGCAACGCTTCGCGCATCTTACGAGGCGCATCAAGAGCGCGCATCAACGATCTCTGCT 2183
2183
2076 CTCGAACGCTTCCAGATCTTACGAGGCGCATCAAGTGGCGGAGATCAAGAGCTCTCTCT 2135
2135
2184 CAAGATGCTGCGCTA 2198
2198
2136 CAAGATGCTGCTTA 2150
2150
RESULT 8
JS-10-374-366-3
Sequence 3, Application US/10374366
Publication No. US20040014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT FILING DATE: 2003-02-26
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
JS-10-374-366-3
Query Match 49.5%; Score 1196.6; DB 16; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2,4e-282;
Matches 1665; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
2Y 99 CGCGCGTCCGACCAAGTGGGCTGCGGCGCCCACTTGCGGCTTCTCGCGGACGCGCGGCT 158
45 CGTGCATCCGCAAGGACGAGCTGTCAATGCGGCTTCAACCACTCGAGTCCGAGGCTT 104
2Y 139 CGATGCGGACCGCGGCGGACGACGATGCGGAGATGCTGAGAGGCTTCTCAAGCGAC 218
105 GCACCTGCCCAACCCAGGTCACGAGGTGACATGCTGAGAGATGCTGCGCGGCT 164
219 CACGACGACGCTGCTGAGCTCAGCGGCTACGCTCACGCTCCGTGAGCTTGTGCGCG 278
165 GACCGAATCGAGCTTGAATCTGACGCTACCTGCTCACTCGAGACGCTGCTGCGG 224
279 CGCGCGCAAGGCGGCGGAGGCTCGGCTGCGGACGAGAGAGATGCTGCGCGGCTGCG 338
225 CGCGAGGAGGCGGAGGCTGCTGCGGCTGAGAGAGAGAGATGCTGCGGAGATGGA 284
339 CAAGAGCGGCGCTTCTCAAGAGGCGGAGCTTCAAGATCGGCTGACGAGTCAACGAG 398
285 CAATATGCTGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
399 TTTGCGTGGCTGCGGCGGACGAGACTGAGAGTCACTGAGCTTCAAGAGGCGCTCAT 458
345 ATTGGCGGATCCGAGACACCGGACGAGAGCGCATCTCGCTCCAGAGGCTTCTCT 404
459 CGAGCACGAGCTTCTGCGGCTGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 518
405 CGAGCACGAGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
519 CGCGAGACGAGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578
465 TCTCGAGAACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 524

579 GCTCAGCGGCGGACGCTGCGGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
525 CTGACCGCGGCGGACGCTGCGGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
639 GAACCAACGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 698
585 CAACCAACGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 644
699 CAGCGGCTTCTGCTGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 758
645 CT 704
759 TTTGCAAGAGGAGGAGGAGATGATGTTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 818
705 GCTTCAAGAGGAGGAGGAGATGATGTTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 764
819 GAGAGGAGTGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 878
765 CGAGCGGCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 824
879 CGCGTGAAGGAGGAGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 938
825 AGCATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 884
939 CTGAGCGGCTTCTCAAGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 998
885 GCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 944
999 CCAAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1058
945 TCAGAGGTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1004
1059 CCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
1005 CTTGAGAGGAGGAGGAGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1064
1119 GGGATTTCTTGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1178
1065 GGGATTTCTTGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1124
1179 CGTGAAGAGTGAATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1235
1125 GCTGAGGAGCTTCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1184
1236 CGACACCGGCTTCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1295
1185 CGACACCGGCTTCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1244
1236 GTCGCTGTCTGATTTGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1355
1245 TGCCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1304
1356 CAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1415
1305 CAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1364
1416 CTTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1475
1365 CTTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1424
1476 TGTTCAGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1535
1425 GCGCTGACGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1484
1536 AGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1595
1485 TGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1544
1596 GAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1655
1545 GTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1604

QY 1556 CGACCTCCGCGCGATGAGAGCTGACTTCAAGAGAGGTTGACCCGCTTCCCGAGCTCT 1715
 DB 1605 CGACTTGGCGGCGACCGAGGTTGAGTTCAAGAGAGGTTGCGCCGACGATGCTGCT 1664
 QY 1716 CTTCCAGACGACCTCGGCACTGGCTCGACGCTCA-----ACGCACTTGGGCT 1763
 DB 1665 CATCGACAGCACTTGGCTCGGCACTGACCGGCTCGAAGCTGGCGGACGAGCTGCGA 1724
 QY 1764 CGAGGTCAAGAGGCGGCTCAAGAGGTTGAGAGAGAGAGAGTACGACTTGGAGCC 1823
 DB 1725 GAAAGTGAACAGAGGCTTGGCCAGAGGCTTGAAGAGAGCAACTGTAAGAGCTGCTCC 1784
 QY 1824 GCGCTGAGACGAGGCTTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1883
 DB 1785 GCGCTGAGACGAGGCTTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
 QY 1884 GCGCTGAGACGAGGCTTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
 DB 1845 GTC-----GCTCTGCTGCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1895
 QY 1944 GCGCATCTGCTCAAG 2003
 DB 1896 GCGCATCTGCTCAAG 1955
 QY 2004 GCGGAG 2063
 DB 1956 GCGGAG 2015
 QY 2064 GCTGAG 2123
 DB 2016 GCTGAG 2075
 QY 2124 GAGCAAGCTCTGCGCATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2183
 DB 2076 CTGCAAGCTCTCAAG 2135
 QY 2184 CAAG 2198
 DB 2136 CAAG 2150

RESULT 9

US-09-939-408A-18
 ; Sequence 18, Application US/09939408A
 ; Patent No. US20020102712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yoshida, Roberta
 ; APPLICANT: Kocistara, Anna
 ; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 ; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 ; FILE OF INVENTION: Using Same
 ; FILE REFERENCE: 29479/500NSCA
 ; CURRENT APPLICATION NUMBER: US/09/939,408A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 09/624,693
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: PCT/US01/23270
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Patent Ver. 2.0
 ; SEQ ID NO 18
 ; LENGTH: 2439
 ; TYPE: DNA
 ; ORGANISM: Rhodotorula toruloides
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2148)
 ; US-09-939-408A-18

Query Match 49.5%; Score 1196.6; DB 9; Length 2439;
 Best Local Similarity: 74.0%; Pred. No. 2.4e-282;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 GCGCGCTCCGACCAAGTGGCTGCGGCGCCCACTTGGCTCTCTCCGCGACGCCCGGCT 158
 DB 45 CGTCCATCCGCAAG 104
 QY 159 CGATGGCCAGCGCGGCAAG 218
 DB 105 GCACTGCGCCCAAG 164
 QY 219 CACTGACAGAGTGTGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 278
 DB 165 GACGAGTCAAGAGTCAAG 224
 QY 279 GCGCGGCAAG 338
 DB 225 GCGAG 284
 QY 339 CAAGAGAGTCAAGTCTCAAG 398
 DB 285 CAAGAGAGTCAAGTCTCAAG 344
 QY 399 TTTGAGAGAGTGGGCGGCAAG 458
 DB 345 ATTGGAG 404
 QY 459 GAGAGAGAGTGGGCGGAG 518
 DB 405 GAGAGAGAGTGGGCGGAG 464
 QY 519 CTTGAGAGAGAGTGGGCGGAG 578
 DB 465 TCTGAGAGAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
 QY 579 GCTGAGAGAGAGTGGGCGGAG 638
 DB 525 CTTGAGAGAGAGTGGGCGGAG 584
 QY 639 GAGAGAGAGTGGGCGGAG 698
 DB 585 CAAGAGAGAGTGGGCGGAG 644
 QY 699 GAGAGAGAGTGGGCGGAG 758
 DB 645 CTTGAGAGAGTGGGCGGAG 704
 QY 759 TTTGAGAGAGAGTGGGCGGAG 818
 DB 705 GCTGAGAGAGAGTGGGCGGAG 764
 QY 819 GAGAGAGAGTGGGCGGAG 878
 DB 765 GAGAGAGAGTGGGCGGAG 824
 QY 879 GAGAGAGAGTGGGCGGAG 938
 DB 825 GAGAGAGAGTGGGCGGAG 884
 QY 939 CTTGAGAGAGTGGGCGGAG 998
 DB 885 GCTGAGAGAGTGGGCGGAG 944
 QY 999 GAGAGAGAGTGGGCGGAG 1058
 DB 945 TCAAGAGAGTGGGCGGAG 1004
 QY 1059 CTTTGGAGAGTGGGCGGAG 1118
 DB 1005 CTTGAGAGAGTGGGCGGAG 1064
 QY 1119 GAGAGAGTGGGCGGAG 1178
 DB 1065 GAGAGAGTGGGCGGAG 1124
 QY 1179 CTTGAGAGAGTGGGCGGAG 1235

1125 CGTAAAGACCTATTACGCCCAACCGCCCTCTACCATTCAGAGCGCGGACAGTCCAGCAGC 1184
1236 CGACACCCCGCTCTCGACGTCAAGAACAGACCGCGCAAGGGGCACTTCCAGGC 1295
1185 CGAACACCTCTCATCAGACGTCAAGAACAGACTTGCACACAGGGGCAATTTCCAGGC 1244
1296 GTGGCGTGTCTCGATTTCAGTAGAAGAACAGAGCTCGACCTGCGCTCATCGCAAGCT 1355
1245 TGCCTGTGTGCGCAACACCATGAGAGAGACTCGCTCGGGCTGCCAATGGCAGAGT 1304
1356 CAACCTTACGAGTGCACCGAGTGTCTCAACGCTGCCATGAACCGCGCTTGCCTTCTGTG 1415
1305 CAACCTTACGAGTGCACCGAGTGTCTCAACGCGCGCATGAACCGCGCTTGCCTTCTGTG 1364
1416 CCTGCGTCCGAGAGACCGGTGCGCTCAACTATCAACGCGAGGGCTTGAATTCATCAATGCG 1475
1365 CCGCGCGCGCAAGACCGCTGCTCTCTTACCACTGCAAGGGCTTGAATTCATCAATGCGCTGC 1424
1476 TGCTTACGCTTCGAGTCCGCGCACCTTGCACACCGCGTCACTTGCCTGCTCAAGCGCGC 1535
1425 GCGGTACACCTCGAGAGTGGAGACCTCGCAACCGCTGTGACAGAGATGTCCAGCGCGC 1484
1536 AGAATGCGACCAACAGCGCGCTCAACTGCTGCTCTCATCTCCGCGCGCGCACTGCGCA 1595
1485 TGAATGCGACCAACAGCGCGCTCAACTGCTGCTCATCTGCGCTGTGCAAGACCGCA 1544
1596 GCGCAAGAGAGCTCTTCTCTCTCTCTGCTGCGCTGCACTTGAACGCTCGAGCGCGCT 1655
1545 GTCCACACAGCTCTTCTCTCTCTCTCTGCTGCGCACTTGAACGCTCTGCAAGCGCT 1604
1656 CGACCTCCGCGCAAGTGAAGTCACTTCAAGAACAGTTCGACCGCGTTCCTCCGACTCT 1715
1605 COACTTGGCGCGATCGAGTTCGAGTTCAGAACAGTTCGCGCCACGCTCTCTCTGCT 1664
1716 CCTCGAGAGAGACCTCGCGCACTGCGCTCGAGCTA-----ACGCACTTGGCGCT 1763
1665 CATGACACAGACTTGTGCTCGCGCACTGAGCTGCTGCAACCTGCGCGAGAGCTCTGCA 1724
1764 CGAGGTCAAGAGGCGCTCAACAGAGCTCTGAGACAGACAGACGATTCGACCTCGAGCC 1823
1725 GAGGTGAACAGAGGCTCGCGCAAGCGCTTGAAGACAGACGATTCGACCTCGAGCC 1784
1824 GCGGTGCAAGAGGCTCTGCTGCTGAGACCGCGCACTGCTGCTCTCTCTCTCTCTCTCT 1883
1785 GCGGTGCAAGAGGCTCTGCTGCTGAGACCGCGCACTGCTGCTCTCTCTCTCTCTCTCT 1844
1884 GCGGTGCAAGAGGCTCTGCTGCTGAGACCGCGCACTGCTGCTCTCTCTCTCTCTCTCT 1943
1845 GTC-----GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
1944 GGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2003
1896 GGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1955
2004 GCGGTGCAAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063
1956 GCGGTGCAAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015
2064 GCGGTGCAAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2123
2016 GCTTGGCGTCAAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2075
2124 GAGCAACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2183
2076 CTCGAACGCTCTCAAGATCTACGAGGCTCAAGTGGGCAAGATCAACACGCTCTCTCT 2135
2184 CAAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2150
2136 CAAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

RESULT 10

US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: B01009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627, 216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147, 719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-7

Query Match 49.4%; Score 1195; DB 9; Length 2151;

Best Local Similarity 73.9%; Pred. No. 5.9e-282; Mismatches 526; Indels 24; Gaps 3;

Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;
99 CGCCGCTCCGACCAAGTGGGCTGCGGCGCCGCACTTGGCTTTCGCGCGACGCGCGCT 158
45 CGTGCATCCGCAAGAGAGGCTGTCATATGCGCGCTCGACCACTTCGAGTGCAGAGCTC 104
159 CGATGCGACGCGCGCGCGACCAAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 218
105 GCACCTGCCCAACCAACGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAG 164
219 CACCGACGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 278
165 GACCGATCCGACGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTC 224
279 CGCCGCAAGGCGCGAGGCTGCGGCTGCGGACCAAGAGTTCGAGTTCGAGTTCGAGTTCGAG 338
225 CGCAGAGAGGCGCGAGGCTGCGGCTGCGGACCAAGAGTTCGAGTTCGAGTTCGAGTTCGAG 284
339 CAAGAGGCTGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAG 398
285 CAATGCGTTCAGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAG 344
399 TTTCGAGTTCGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAG 458
345 ATTTCGAGTTCGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAG 404
459 CGAGCACAGTTCGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAG 518
405 CGAGCACAGTTCGAGTTCGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAG 464
519 CCTGAGAACAGGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 578
465 TTTCGAGAACAGGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 524
579 GCTCAAGGTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 638
525 CTTGACCGCGCGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
639 GAACACCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCAT 698
585 CAACACCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCATCAAGCGCAT 644
699 CAGCCGCTCTGTCATCAAGCGCGCGCGCATCAAGCGCGCATCAAGCGCGCATCAAGCGCGCAT 758
645 GTCCT 704
759 TTTCGACGAGGACCAAGGATCAATGTTGCGCGCGAGGCACTTCGCTCTTGGGCT 816

Db 705 CGTCCAGAGGAGGAGAGAGATCTGTACGCCCCGAGGAGGATGGCGCTTTTACCT 764
 Qy 819 CGAGGAGATGTCCTCGCCCGAAGAGAGGTCTCGATCTGTCAACGAGAGCGCGTCTC 878
 Db 765 CGAGCCCGTCTCTCGGCCCGAAGAGAGTCTCGGTCTCGTCAACGAGACCGCGTCTC 824
 Qy 879 CGCCTGATGAGGAGACCTCAGTCTGACAGTCTGACATGCTCTGCTCTCTGAGGC 938
 Db 825 AGCATGATGACACCTTCTGCTCTGACAGCTCAATGCTCTGCTCTCTGACAGTC 884
 Qy 939 CTTGACGAGCTCTCAAGGTGAGAGGCAATGATCGAGGAGAGGCTTCTGAGCGCTTCAT 998
 Db 885 GCTCAGCGCCATGACGATGAGAGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 944
 Qy 999 CAGCAGAGCTGTCGCGCCGAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1058
 Db 945 TCACGAGTACGCGCCCTCACCCGAGCAGATCAATGCTCGGAGAAACATCCGCAAGCT 1004
 Qy 1059 CCTTTCGAGCTGTCGCTTGGCTTGGAGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1118
 Db 1005 CTTGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1064
 Qy 1119 GAGGATTTCTTGGCCAGAGAGCGCTACCGCTCCGACAGTCCGCTCAGTTCTCGCGCGCT 1178
 Db 1065 GGGCATTTCTCCGCGAGAGCGCGCTACCGCTCCGACAGTCTCAGTGGCTCGGCGCGCT 1124
 Qy 1179 CGTGGAGAGCATGATGACAGGCTTACGATCTCTCGCTGAGAGAACAA---GAGAGAC 1235
 Db 1125 CTTGAGGAGAGCTTCAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1184
 Qy 1236 CGAACAACCGCTCTCTGAGAGTCTGAGAGAACAGAGAGAGGAGGAGGAGGAGGAGGAGG 1295
 Db 1185 CGAACAACCTCTCATGAGAGTCTGAGAGAACAGATTTCCAGAGGAGGAGGAGGAGGAGG 1244
 Qy 1296 GTCCGCTGTCTGATTTGATGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1355
 Db 1245 TCCGCTGTGAGCAACACATGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1304
 Qy 1356 CAATTTCAAGAGTGTGACAGGAGTGTCAACGCTGAGGAGGAGGAGGAGGAGGAGGAGG 1415
 Db 1305 CAATTTCAAGAGTGTGACAGGAGTGTCAACGCTGAGGAGGAGGAGGAGGAGGAGGAGG 1364
 Qy 1416 CTTGCTGCTCGAGAGAGCGCTGTCAACATGACAGGAGGAGGAGGAGGAGGAGGAGG 1475
 Db 1365 CTTGCTGCTCGAGAGAGCGCTGTCTCTCACTGAGAGGAGGAGGAGGAGGAGGAGGAGG 1424
 Qy 1476 TGCTTACGCTTCCGAGAGTCCGCGACCTTGGCAACCGGAGGAGGAGGAGGAGGAGGAGG 1535
 Db 1425 GAGGAGAGCTTCCGAGAGTGTGAGAGAGCTTGGCAACCGGAGGAGGAGGAGGAGGAGG 1484
 Qy 1536 AGAGATGAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1595
 Db 1485 TGAAGATGAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1544
 Qy 1596 GAGGAGAGAGCTCTTCT 1655
 Db 1545 GTCAAGAGAGCTCTTCT 1604
 Qy 1656 CGACCTCGCGAGATGAGAGTGTCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1715
 Db 1605 CGACCTTGGCGAGATGAGAGTGTCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1664
 Qy 1716 CTTGAGAGAGAGCTTGGAGAGTGTCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1763
 Db 1665 CATGAGAGAGAGCTTGGAGAGTGTCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1724
 Qy 1764 CGAGGATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1823
 Db 1725 GAGAGTGAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1784
 Qy 1824 GCGCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1883

Db 1785 GCGCTGAGCAGAGCGCTTCTCTTGGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1844
 Qy 1884 GCCCTGTGCAACAGTCAACCTTACTGCGGTCAACGCGGTGAGAGGAGGAGGAGGAGGAGGAG 1943
 Db 1845 GTC-----GCTCTGCTGCGCGCGGTCAACGCGGTGAGAGGAGGAGGAGGAGGAGGAGG 1895
 Qy 1944 GGCATCTTCTGCTCAAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2003
 Db 1896 GGCATCTTCTGCTCAAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1955
 Qy 2004 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2063
 Db 1956 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2015
 Qy 2064 GCTGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2123
 Db 2016 GCTTGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2075
 Qy 2124 GAGGAGGAGTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2183
 Db 2076 CTGAGAGGAGTCTCAAGAGTCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2135
 Qy 2184 CAAGATGCTCGAGTCA 2198
 Db 2136 CAAGATGCTCGAGTCA 2150

RESULT 11
 US-10-188-523B-7
 ; Sequence 7, Application US/10188523B
 ; Publication No. US20030170834A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-Song
 ; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
 ; FILE REFERENCE: BCI1009 US DIVCIP
 ; CURRENT APPLICATION NUMBER: US/10/188,523B
 ; CURRENT FILING DATE: 2002-10-17
 ; PRIOR APPLICATION NUMBER: US 09/627,216
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR APPLICATION NUMBER: US 60/147,719
 ; PRIOR FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 7
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: Rhodotorula glutinis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2151)
 ; OTHER INFORMATION:
 US-10-188-523B-7

Query Match 49.4%; Score 1195; DB 15; Length 2151;
 Best Local Similarity 73.9%; Pred. No. 5,9e-282;
 Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

Qy 99 CGCGCTCCGAGCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 158
 Db 45 GTTGGATCCGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 104
 Qy 159 CGATGAGCAGCGCGGAGCAGCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 218
 Db 105 GCACCTGCGCAACAGCAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 164
 Qy 219 CAGCGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 278
 Db 165 GACGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 224
 Qy 279 CGCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 338
 Db 225 CGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 284

2Y	339	CAAGAGCGTGCACCTTCTCTCAAGGCCAGAGCTGTAGAACTCGGCTTACGGAGTCAACAAGG	338
2Y	286	CAATCGGTGCAATTTCTTGGCTTGCAACTTCTCATAGACGCTTTCAGCGCTCAAGACTCG	344
2Y	339	TTTCGTGGCTCGCGCGACACAGAGACTGAGATGACGATCAAGCCCTCCAGAAAGGCGTTCAT	458
2Y	345	ATTTCGCGGATCCCGACACACCGCACCGAGAGCGGCATCTCGCTCCAGAAAGGCGTCTCT	404
2Y	459	CGAGACACCAAGTCTTGGCGGCGTGAACGCCGAGAGTCCGCTGTCTTCAAGGCTGAGACCGG	518
2Y	405	CGAGACCAAGTCTTGGCGGCGTGAACGCCGAGAGTCCGCTGTCTTCAAGGCTGAGACCGG	464
2Y	519	CCTCGAGAACACGCGTTCCGCTCGAGAGTGTCTCCGCGCGGCATATGTCATCCGCTCAATC	578
2Y	465	TCTGAGAACTCGGTTCTCCCTCCGAGGTTGTTGGCGCGCATGACATACCGGCTCAACAG	524
2Y	579	GCTCAACGCTGGCCACTTCGCGCGTTCGCGCTGTCTTGAAGCGCTCAACCACTTCT	638
2Y	535	CTTGACCCCGCGGCACTCGGCTGTCCGCGCTGTCTTGAAGCGCTCAACCACTTCT	584
2Y	639	GAACACCGCATCAACGCGCATATCGCCCGCGCGGCTCAATCTCGGCGTCCGCGGCACT	698
2Y	585	CAACACCGCATCAACCGCATATCGCCCGCGCGGCTCAATCTCTGGGTGGGCGAAGCT	644
2Y	699	CAGCCGCTCTTGATCATTCGCGCGGCATCAACGCTCAACCCGACGTTAAAGTTCAAGT	758
2Y	645	GTCCTCTCTCTCAATTCGAGCGGCATCAACGCTTCAACCGGACAGAGTGTCAAGT	704
2Y	759	TTTGGACGAGGGAACCGAAGAAATCAATGTTTGGCGCGAGGCAATCGGCTTTTGGCT	818
2Y	705	CGTCCACAGAGGCAAGAGAGAAATCTGTATGCGCCCGAGGCGAGATGGCGCTTTCAACT	764
2Y	819	CGAGCAGTCTCTCGCGCGCGGAGAGAGGATCTCGGCTTGTCAACGGAACCGCGGCTC	878
2Y	765	CGAGCCGCTGTCTCGCGCGCGGAGAGAGGATCTCGGCTTGTCAACGGAACCGCGGCTC	824
2Y	879	CGCCTCATATGCGGACCTCTCACTGTGACAGCATGCGCATATGCTCGCTCTCGGACGCG	938
2Y	825	AGCATCGATGACGACCTCTCGCTGTGACAGCGTCAATATGCTCTCGCTCTTTCGACATC	884
2Y	939	CTTGAACGCTCTCAACGCGTGAAGGACATATGTCGCGCACAGAGGCTGTGTCGCGCTTCAT	998
2Y	885	GCTTACGCGCATTAACGCTGGAAGGATATGTCGCGCACAGCGGCTGTGTCGCGCTTCAT	944
2Y	999	CCAGACGCTTTCGCGCGCGCACCGCGCGGACGCTGTGAGAGTTCGCGCGCAATCTCGACGCT	1058
2Y	945	TACACAGCTACGCGCGCTTCACTCCGACGAGATCGAAGTGGCGGAAATCTCCGAACT	1004
2Y	1059	CCCTTTCGCGCTGTGCTTTCGCTTGAAGCAGAGAGAGTCAAGTCAAGTCAAGAGAGAG	1118
2Y	1005	CCTCGAGGGAACCGCTTTCGCTGTTCACATTAAGAGAGTCAAGTCAAGAGAGAG	1064
2Y	1119	GGGCAATCTTTCGCGAGACCGGCTACCGGCTCGGACGTCGCGCTCAATTCCTCGCGCGCT	1178
2Y	1065	GGGCAATCTTTCGCGCGAGACCGGCTACCGGCTTTCGCGACGCTCTCAATTCCTCGCGCGCT	1124
2Y	1179	CGTGAAGACATTAAGACAGCGCTTCACTGATCTCTGCTCGGAAACAACA---CGACGAC	1235
2Y	1125	CGTGAAGACATTAAGACAGCGCTTCACTGATCTCTGCTCGGAAACAACA---CGACGAC	1184
2Y	1236	CGACACCGGCTCTTCGACGCTCGAGAACAGACAGACCGGCGACCGCGGCATTCGAGC	1295
2Y	1185	CGACACCGGCTCTTCGACGCTCGAGAACAGACGCTCGGCGGCATTCGAGC	1244
2Y	1296	GTGCGGTGTCTGATTTGATGAGAGAACACGAGCTGCACTGCGCTTATCGGCACACT	1355
2Y	1245	TGCGGTGTGAGCAACACATGAGAGAACGCTCGGCGGTGCGGCAGATGCGCAACT	1304
2Y	1356	CAACTTCAAGAGTGCACGAGTTGCTCAACGCTGACATAACCGCGGCTGCTCTTCTG	1415
2Y	1305	CAACTTCAAGAGTGCACGAGTTGCTCAACGCTGACATAACCGCGGCTGCTCTTCTG	1364
2Y	1416	CGTGTGCGAGAACCGGTGCTTCAACTATCAAGCGCAGAGGCTTGAACATTCATCATGCG	1475

Db	1365	CCTGCGCGCGAAGACCCTTCGTCTCTCTCAACCACTGCAAGAGGCGCTCGACATCGCCGCTGC	1424
Qy	1476	TGCTTAAAGCTTTGGAGAGCTCGGSCCACTTTGSCCAACCCGGGTCACTTACTTCTGTCCAGCCCGC	1535
Db	1425	GCCGTAACCTCTGGAGATTGGGAGAACCTTCGCAACCTGTGAGAGACGCAATGTCCAGCGGC	1484
Qy	1536	AGAGATGGGCAACACGAGCCGCTCAACTCGCTCGCTCATCTTCGCGCCGCCACATGCGGA	1595
Db	1485	TGAGATGGCGAACCAGCGGGGTCAAACTCGTTGGCGCTCATCTCGGCTGTGCCAGACGGA	1544
Qy	1596	GCCCAACGACGTCTTCTTCTCTCTCTCTCGCTTCGACCTGTACTGACAGCTTCAGGCGCT	1655
Db	1545	GTCCAAACGATCTTCTTCTCTCTCTCTCGCAACCACTCTACCTGCTTCCCAAGCAT	1604
Qy	1656	CGACCTCGCGCGATGGAGGCTCGACTTCAAGAAAGAGTTGCAACCGCGCTTCTCCGACCT	1715
Db	1605	CGACTTGGCGCGATCGAGTTTCAGTTTCAAGAAAGAGTTTGGCCCAAGCATGTCTCGCT	1664
Qy	1716	CCTCCAGACGACCTCGGCACTGAGCTTCAAGTCA-----ACGACTTGGCGCT	1763
Db	1665	CATCCACGACGACTTTTGGCTTCGCGCATGACCGGCTCGAACTTGCAGCGACGAGCTGTGGA	1724
Qy	1764	CGAGGTCAAGAAAGGGGCTCAACAAGCGTTCGAGACGACGACGATCGATCGACCTCGACC	1823
Db	1725	GAAAGTAAACAAGACGCTCGCCAAAGCGCTTCGAGACGACCACTGTAGACTGTGTCC	1784
Qy	1824	GCGCTGGACGACGCGCTTCTGTGACCGGACCGGACCGTGTGAGCTTCTGTGCTCTC	1883
Db	1785	GCGGTGGACGACGCGCTTCTGTGCGCGCGGCAACGTCGTGAGAGTCTTCTGTGAC	1844
Qy	1884	GCCCTTGCACAGCTTCACTTACTGCGGTGCAACGCGTGAAGGTTGCGTCCGCGCGAAGA	1943
Db	1845	GTCTGTGCTGTGCGCGCGGTCAACGCTTGAAGGTGGCGCGCGCAATC	1895
Qy	1944	GCCCATCTCGCTCAACGCGCGAGGTGCGCAACCGCTTCTGCGAAGCGCGTTCGCAAGC	2003
Db	1896	GCGCATCTTGCCTCAACCGCGCAAGTCCGCGAGACTTCTGTGTCCGCGGTGACCTCTGC	1955
Qy	2004	GCGCGGCAACGATTACTTCTGTGCGCGCAACGCGCGTCTGTACTCGTTGTGTGCGGAGGA	2063
Db	1956	GCCCGGCTCTGTGATCTCTGTGCGCGCACTCGAATCTCTTACGCGCTTGTGTGCGAAGA	2015
Qy	2064	GCTCGGCGGTGACGAGCGCGCGCGGCGACGTGTGTGTGCGCGTGCAGAGACGATCGG	2123
Db	2016	GCTTGGCGGTCAAGCGCGCGCGGAGACGTCTTCTGTGCGAAGAGATGACGATCGG	2075
Qy	2124	GAGCAACGTTCTGCGCATTCAGAGGCGATCAAGAGCGCGCATCAACAGTCTCTCT	2183
Db	2076	CTCGAAGCTTCTCAAGATCTTCAGAGGCGATCAAGTGGGAGAGATCAACAGTCTCTCT	2135
Qy	2184	CAAGATGCTCGCGTA	2198
Db	2136	CAAGATGCTCGCTTA	2150
RESULT 12			
US-10-374-366-1			
Sequence 1, Application US/10374366			
Publication No. US20040014085A1			
GENERAL INFORMATION:			
APPLICANT: Tang, Xiao-SONG			
APPLICANT: Milano, Joseph			
FILE REFERENCE: CL1194 US NA			
CURRENT APPLICATION NUMBER: US/10/374,366			
CURRENT FILING DATE: 2003-02-26			
PRIOR APPLICATION NUMBER: 60/360,279			
PRIOR FILING DATE: 2002-02-26			
NUMBER OF SEQ ID NOS: 203			
SOFTWARE: PatentIn version 3.2			
SEQ ID NO 1			
LENGTH: 2151			

2076 CTGCAACGCTCCAGATCTTACAGAGGCATCAAGTCGGCGAGAGATCAACAGCTCTCTCT 2135
2184 CAAGATGCTCGCGTA 2198
2136 CAAGATGCTCGCTTA 2150

RESULT 13
JS-10-374-366-9
Sequence 9, Application US/10374366
Publication No. US20040014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/360,279
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
JS-10-374-366-9

Query Match 49.4%; Score 1195; DB 16; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5,9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

99 CGCGCTTCGACCAAGTCGGCTCGGGCCCCACTTGGCTCTCCGCCGACCGCGGCT 158
Db CGTGCATCCGCAAGGAGGCTGTCAATGGCGCTTCACCACTCGACAGTCGACGGCTC 104
2159 CGATGGCGACGCGCGGACGACGATGCGAGTCGATGTCAGAGAGTCCTCGACGCC 218
Db GCACCTGCCCAACCAAGGTCACGCGAGTCGATGTCGAAAGATGTCGCGCGCTC 164
219 CACCGACGACCTCGTCGAGCTCAAGCGGTACAGCTCAGCTCGTTCGCGCTC 278
Db GACCGACGACGCTCGCAAGCTCGAGGCTACGCTCAGCTCGTTCGCGCTC 224
2165 GACCGACGACGCTCGCAAGCTCGAGGCTACGCTCAGCTCGTTCGCGCTC 224
279 CGCGCGCAAGGGGCGAGGTCGCGCTTCGCAAGCAGAGCAGATCCGCGACGCTCGA 338
Db CGCAGAGAGGCGAGGCTCGCTCGCTCAAGAGCAGAGCAGATCCGCTCAAGATTA 284
225 CGCAGAGAGGCGAGGCTCGCTCGCTCAAGAGCAGAGCAGATCCGCTCAAGATTA 284
339 CAAAGAGCTGACTTCTCTCAAGGCTTCAAGCTTCAAGCTTCAAGAGTCAACAGCG 398
Db CAAATCGCTGAGTTCTTGGCTCGCAACTCTCATAGAGTCTTCAAGGCTTCAAGCTG 344
399 TTTGCTGCTCGCGCGACACGAGAGTGAAGATGAGTCAAGCTTCAAGAGGCTCAT 458
Db ATTGGCGAGATCCGAGAGACCCGACCGCAGAGAGCAGATCCGCTCAAGAGCTTCT 404
459 CGAGAGCAGCTGCGCGCGTGAAGCGAGCTCGCTCGCTTCAAGCTTCAAGAGCGCG 518
Db CGAGAGCAGCTGCGCGGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
405 CGAGAGCAGCTGCGCGGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
519 CCTCGAGAGCAGCTTCT 578
Db TCTCGAGAGCAGCTTCT 524
465 TCTCGAGAGCAGCTTCT 524
579 GCTTACGCTGCGCT 638
Db CTTGACCGCGCGCGCT 584
535 GAACACCGCATCAAGCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 698
Db CAACACCGCATCAAGCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 644
639 CAGCCGCT 758

Db 645 CT 704
Qy 759 TTTGCAAGAGGAGACCGAAGATCATGTTTGGCGCGAGGCGCATCTGCTTTGGTCT 818
Db 705 GCTTCAAGAGGAGCAAGAGATCTCTTACCGCCGCGAGGCGAGTGGCTTCTTCAACT 764
Qy 819 CGAGGAGAGTCT 878
Db 765 CGAGCCGCTGCT 824
Qy 879 CGCTCGATGGCAGCCCTGAGTCTGACGAGTCTGCAATGCTCTCTCTCTCTCTCTCT 938
Db 825 AGCATGATGGCAGCCCTGAGTCTGACGAGTCTGCAATGCTCTCTCTCTCTCTCTCT 884
Qy 939 CTTGAGGCTCTTCAAGTGGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCT 998
Db 885 GCTCAGGCGCATACGATGAGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCT 944
Qy 999 CCAAGAGTCT 1058
Db 945 TCAAGAGTCAAGCGCGCTTCAAGCGAGTCAAGTCTGAGGAGTCAAGAGTCAAGAG 1004
Qy 1059 CTTTCCGCT 1118
Db 1005 CTTGAGGAGAGCGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1064
Qy 1119 GGGGATTTCTTCCGCAAGACCGCTTACCGGCTCTGAGAGTCTGAGGAGTCTCTCT 1178
Db 1065 GGGGATTTCTTCCGCAAGACCGCTTACCGGCTCTGAGAGTCTGAGGAGTCTCTCT 1124
Qy 1179 CGTGAAGAGATGATGACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
Db 1125 CTTGAGGAGAGCGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1184
Qy 1236 CGACCAAGCGCT 1295
Db 1185 CGACCAAGCGCT 1244
Qy 1296 GTCGCTGCT 1355
Db 1245 TGCCTGCTGCT 1304
Qy 1356 CAACTTACCGAGTCAAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1415
Db 1305 CAACTTACCGAGTCAAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1364
Qy 1416 CTTGCTGCGAGAGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1475
Db 1365 CTTGCTGCGAGAGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
Qy 1476 TGTCTTACGCTTCT 1535
Db 1425 GGTCTTACGCTTCT 1484
Qy 1536 AGAGATGGGCAACCGAGCGTCAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
Db 1485 TGAATGGGCAACCGAGCGTCAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1544
Qy 1596 GGGCAAGAGCT 1655
Db 1545 GTCGAGAGAGCT 1604
Qy 1656 CGACCTCGCGCGATGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1715
Db 1605 CGACTTGGCGCGAGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1664
Qy 1716 CTTGAGAGAGACT 1763
Db 1665 CATGAGAGAGACT 1724
Qy 1764 CGAGGTCAAGAGGCGCTCAAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1823

Db 1725 GAAGTGAACAAGAGCTCGCAAGCGCTCGAGCAGACCAACTCGTACGACTCGTCCC 1784
QY 1824 GCGCTGACGACGAGCTTCTCGTACGACGACGACGACGCTCGTACGCTCTGCTC 1883
Db 1785 GCGCTGACGACGAGCTTCTCGTACGACGACGACGACGCTCGTACGCTCTGCTC 1844
QY 1884 GCGCTGACGACGAGCTTCTCGTACGACGACGACGACGCTCGTACGCTCTGCTC 1943
Db 1845 GTC-----GCTCTGCTCGCGCGCTCGACGCTCGAGAGTCCCGCGCGAGTC 1895
QY 1944 GGCATCTCTCGTACGACGACGACGACGACGCTCGTACGACGACGCTCGACGAC 2003
Db 1896 GGCATCTCTCGTACGACGACGACGACGACGCTCGTACGACGACGCTCGACGAC 1955
QY 2004 GCGCGGACGACGACGCTCTGCTCGCGGACGACGCGGCTCTGCTCTGCTCGGACG 2063
Db 1956 GCGCGGCTCTGCTCTGCTCTGCTCGCGGACGACGCTCTGCTCTGCTCGGACG 2015
QY 2064 GCTCGGCTGACGACGCGCGCGGACGAGTGTGTTGTGGGCTGACGAGACGATCGG 2123
Db 2016 GCTTGGCTCAAGGCGCGCGCGGACGAGTGTGTTGTGGGCTGACGAGACGATCGG 2075
QY 2124 GAGCAAGCTCTGCGCATCTACGAGGCGCATCAAGACGCGGATCAACGCTCTCT 2183
Db 2076 CTGCAAGCTCTCAAGATCTACGAGGCGCATCAAGTGGGCGAGATCAAGAGCTCTCT 2135
QY 2184 CAAGATGCTCGCTGA 2198
Db 2136 CAAGATGCTCGCTGA 2150

RESULT 14
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodospiridium rubrum
US-10-374-366-13

Query Match 49.4%; Score 1195; DB 16; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5.9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

QY 99 CGCGCTCGACCAAGTGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 158
Db 45 CGTGGCTCGACCAAGTGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 104
QY 159 CGATGCGACGCT 218
Db 105 GCACTTGGCGACCAAGTGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 164
QY 219 CACGACGACGCTCTGAGCTCAAGCGGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 278
Db 165 GACGACGACGCTCTGAGCTCAAGCGGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 224
QY 279 CGCT 338
Db 225 CGCT 284
QY 339 CAGAGGCTGAGCTTCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 398

Db 285 CAAATCGAGTCTTGGCTCGCAACTCTCATAGAGGTCTACGCGGTGACGACTCG 344
QY 399 TTTGGGTGGCTCGGCT 458
Db 345 ATTGGGAGATCGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 404
QY 459 CGAGCAGGCTCTGCT 518
Db 405 CGAGCAGGCTCTGCT 464
QY 519 CTTGAGAAACGCTTTCGCTCGAGGCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCT 578
Db 465 TCTGAGAAACGCTTTCGCTCGAGGCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCT 524
QY 579 GCTCAGCGGCTGCT 638
Db 525 CTTGAGAAACGCTTTCGCTCGAGGCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCT 584
QY 639 GAACCAACGATACGCT 698
Db 585 CAACCAACGATACGCT 644
QY 699 CAGCGCGCTCTGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 758
Db 645 CTTCTCTCTCTCAAGTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 704
QY 759 TTTGACAGAGGAAACGAGAGATCATGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCT 818
Db 705 GGTCAAGAGGAAACGAGAGATCATGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCT 764
QY 819 CAGAGCAGTCTCTGCT 878
Db 825 AGCATGATGCT 884
QY 879 CGCCTGATGCT 938
Db 765 CGAGCGGCTCTGCT 824
QY 939 CTTGACGCTCTGACGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 998
Db 885 GCTCAGGCGATGACGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 944
QY 999 CCAAGAGCTGCT 1058
Db 945 TCAAGAGCTGACGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1004
QY 1059 CTTTCCGCTCTGCTTGGCGGTGAGCAGAGGAGGATCAAGGATCAAGGAGCAGCA 1118
Db 1005 CTTGAGGAGGAGCGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
QY 1119 GGGCATTTCTGCGCAGAGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1178
Db 1065 GGGCATTTCTGCGCAGAGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1124
QY 1179 CGTGAAGACATGAGCAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1125 CGTGAAGACATGAGCAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
QY 1236 CGACCAACCGCTCTGAGTGAAGACAGCAGCGCGCGCGCGCGCGCGCGCGCT 1295
Db 1185 CGACCAACCGCTCTGAGTGAAGACAGCAGCGCGCGCGCGCGCGCGCGCGCT 1244
QY 1296 GTGCGCTGCTGATTTGATGAGAAACAGGCTGCACTCGCGCTCATGCGCAAGCT 1355
Db 1245 TGCCTGTTGGCAACACCATGAGAAAGCTGCGCTCGCGCTCGCGCGCGCGCGCT 1304
QY 1356 CAATTCAAGAGTGAAGTGTCTCAAGCTGCTGCAAGCGCGCGCGCGCGCTGCT 1415
Db 1305 CAATTCAAGAGTGAAGTGTCTCAAGCTGCTGCAAGCGCGCGCGCGCGCTGCT 1364
QY 1416 CTTGCGTGGAGAGCGCGCTGCTCAATTACGCGCGCGCGCGCTGCAATTACATGCT 1475

Db	1365	CCTCGGCGCCGAAGACCCCTTGCTCTCTACCATCTGCAAGGGCTCGACATCGCCGCTGC	1424
2y	1476	TGCTTACCTTTGAGGCTCGGCACTTTGCCAACCCGGTCACTACCTTGTGCCAGCCGC	1535
Db	1425	GGCGCAACCTCTGGAGTTTGGGACACCTTGGCAACCTTGTGACGACGATGTCCAGCGGGC	1484
2y	1536	AGAGATGGGCAACCGGGCGTCACTGGCTGGCTATCTACCGAGCGCGCATCGCCGA	1595
Db	1485	TGAGATGGCGAACGAGCGGTAACTTCGTTTGGCTCATCTTCGGCTCGTCCGACAGCCGA	1544
2y	1536	GGCCCAAGACGTCTTTCTCTCTCTCGCCCTCGACCTGTATCTGACAGCTCAGGCCGT	1655
Db	1545	GTCCAAACGACGCTTTCTCTCTCTCTCTCGCCACCACTTACTGCGTTCTTCCAAGCAT	1604
2y	1656	CGACTCTGGCGGAGTGAAGAGCTGATTGAAGAGAGTTGGCCCAAGCATGTCTGCT	1664
Db	1716	CTCCAGAGACACTCTGGGCACTGGGCTCGACGTCA-----ACGCACTTGGCT	1763
2y	1764	CGAGGTCAAGAAAGCGCTTCAACAAAGCTTCGAGTGAAGACGACGATCGACTTGAACC	1823
Db	1725	GAAAGGTGAACAAGAGGCTTCGACACGCTCGACGACGACAACTGTACGACCTGTCC	1784
2y	1824	GGCTGGGACGACGCTTCTGTGACGACGCGACCGCTCGACGCTCTCTGCTCTC	1883
Db	1785	GGCTGGGACGACGCTTCTGTGACGACGCGACCGCTCGACGCTCTCTGCTCTC	1844
2y	1884	GGCTGTGTGCCAAAGTCAACCTTACTGCTGCTGCTGCAAGCGTGAAGAGTTGCTTGGCCGAGAA	1943
Db	1845	GTC-----GCTCTGCTGCTGGCGCGCTGCAACCGCTGAAGAGTGGCGCGCGCGAGTTC	1895
2y	1944	GGCCATCTGCTCAGCGCGGAGAGTGGCAACCGCTTCTGAGAGAGCGCGTCTTGGCAAGC	2003
Db	1896	GGCCATCTGCTCAGCGCGGAGAGTGGCAACCGCTTCTGAGAGAGCGCGTCTTGGCAAGC	1955
2y	2004	GGCGGCGACGACATCTCTCGCCGCGACGCGCGCTCTGTACTGTTGTTGTCGCGGAGAA	2063
Db	1956	GGCGGCGCTCTGTACTCTCTCGCCGCGACCTCAATCTCTGACGCTTCTGTCTCGGAGAA	2015
2y	2064	GCTCGCGCTGAGAGGCGCGCGGCGAGACGTGTTTCTCGCGCTGACGACGAGAGACGATCGG	2123
Db	2016	GCTTGGCGTCAAGGCGCGCGGAGAGCGTCTTCTCTCGCAAGCAAGAGTGAAGATCGG	2075
2y	2124	GAGCAACGCTCGCCCATCTAGAGAGGATCAAGAGAGCGCGCATCAACCAAGCTCTCTCT	2183
Db	2076	CTCGAAGCTCTCAAGATCTAAGAGGATCAAGAGGATCAAGAGAGATCAACCAAGCTCTCTCT	2135
2y	2184	CAAGATGCTCGGCTA	2198
Db	2136	CAAGATGCTCGGCTA	2150
RESULT 15			
US-10-374-366-23			
Sequence 23, Application US/10374366			
Publication No. US20040014085A1			
GENERAL INFORMATION:			
APPLICANT: Tany, Xiao-Song			
APPLICANT: Milano, Joseph			
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS			
FILE REFERENCE: CL1/94 US NA			
CURRENT APPLICATION NUMBER: US/10/374,366			
PRIOR FILING DATE: 2003-02-26			
PRIOR APPLICATION NUMBER: 60/360,279			
NUMBER OF SEQ ID NOS: 203			
SOFTWARE: PatentIn version 3.2.			
SEQ ID NO 23			
LENGTH: 2151			
TYPE: DNA			

! ORGANISM: Rhodospiridium glutinis
US-10-374-366-23

Query Match	49.4%	Score 1195;	DB 16;	Length 2151;
Best Local Similarity	73.9%	Pred. No. 5.9e-282;		
Matches 1564;	Conservative	1;	Mismatches 526;	Indels 24;
				Gaps 3;

QY	99	GGCGGCTCCGACCAAGGTCGGCTGGGGGCCCCACCTTGCGCTCTCGCGCGCACGGCCGGCTCT	158
Db	45	CGTGCATCCGCGCAAGAGGAGGTGTCAATGAGCGCTGCACCAACTTCGCAATTCGCAAGGCTC	104
QY	159	CGATGGCCACGCGCGCGCACAGTGCAGCTCGAGATCTGACAGAGTCTCTCAGCGACCC	218
Db	105	GCACCTGCTCCCAACAACCAGAGTCAAGCAGGTGCACATGATCGAAGAAAGATGCTCGCGCGCC	164
QY	219	CACCGACGACGTGGTGGAGCTCAAGCGGTACAGCTCACCGTCGATGACGTTGCGCGCGC	278
Db	165	GACCGACTTCAGCTGTCAACTTCAGCGCTACTGTCTAACCTTGGAGAAGTGGTCTTGCGC	224
QY	279	CGCCCGCAAGGGGCGCAGGGTCCGCTCCAGAAAGACGAGAGATCCGCGCACGGCTGCA	338
Db	225	CGCGAGGAAGGGCAGGGCTGTCCGCGCTCAAGGACAGGACGAGATCGCTCAAGATGTGA	284
QY	339	CAGAAGCTGCACCTCTCTCAAGGCCACGCTTCAGAACTCGGTCTACGAGATTCACACGGG	398
Db	285	CAATTCGGTGGAGTTCTTGGCTGCACACTCTCCATGAGCGTCTACGCGCTCAAGATGG	344
QY	399	TTTGGTGGCTCGGCGCGACACGAGAGCTGAGAGATGACGTAGCTCCAGAGGCGCTCAT	458
Db	345	ATTGGGGGAGTCCGCGACACCCGCGACGAGAGAGCCATCTCGTCCAGAAAGGCTTCCT	404
QY	459	CGAGCACAGCTTCGGGCGGTGACGCCGACGTCCGTCGTCTTACGCGTCCGACCGCG	518
Db	405	CGACACACAGCTTCGGGCTGTCTCCCTTGGTTCGTTGACCTGTTCCGCTTCGCGCCGCGG	464
QY	519	CCTCGAGAACAGCTTCGCTCGAGAGTGTCCGCGGCGCCATGGTTCATCCGCGTCACTC	578
Db	465	TCTGGAGACTCGCTTCCCTCGAGGTTTTCGCGGCGCATGACAACTCCGCGTCAACAG	524
QY	579	GCTCACGGTGGCCACTCGGCGGTCCGCTCTGCTGTCCTTGAAGCGCTACCAATTCTT	638
Db	525	CTTACCCGCGGCGCACTGGGTGTCCGCTCTGCTGTGCTCGAGGCGCTCAACAATTCTCT	584
QY	639	GAACCAACGATTCACGGCCATTCGCCCCCTCCGCGGGTCCATCTCCGCGTTCGGGCGACT	698
Db	585	CAACCAACGCACTACCCCATCTGCCCCCTCCGCGGACCACTCTGCTGTCGGGCGACT	644
QY	699	CAGCCGCTCTCGATACTCGCGCGCGCCATCACCGGTACCCCGACGTCAGAGTTACGT	758
Db	645	CTCTCTCTCTCTCAATGCAATGCAAGGGCCATCAGCGGTACCCCGACAGCAAGGTGACGT	704
QY	759	TTTGACGAGGGAACCGAAGATCATGTTTGGCGCGGAGGGCACTCTCGCTTTGGTCT	818
Db	705	CGTCCACGAGGGCAGAGAGATCCGTACGCGCGCGGAGGGCGATGGCGCTTTCAACT	764
QY	819	CGAGGAGATCGTCTCTCGGCGCGAAGAGAGGCTCGTCTGGTTCACAGGAACGACGCTCTC	878
Db	765	CGAGCCGCTCTCTCGGCGCGAAGAGAGGCTCTGCTCAACCGCAACCGCGCTCTC	824
QY	879	CGCTTCAGTGGCAACCTTCAGTCTTGCAAGACTGCAACATGCTCTTCCGCTCTCTTTCGACGC	938
Db	825	AGCATTCAGTGGCCACCTCGCTCTGACACGACGACACATGCTCTCGCTCTCTCGACATC	884
QY	939	CTTACAGGCTTCAAGGTGGAGGCAATGATGGGACAGAGGGCTCGTTCGGCGCTTCAT	998
Db	885	GCTACGGCGCATGAGGTCGAGAGGATGATGGCGACGCGGCTCTGTTCAACCCCTTCTCT	944
QY	999	CCAGAGCTTCCGCGCGCGCAACCCCGGACAGTCCAGGTCGCGGCAACATCCGACGCT	1058
Db	945	TCAAGAGTCAAGCGGCTCTCAACCGACGACATCGAATCCGCGGAAACATCCCAAGCT	1004
QY	1059	CCTTTCGGGCTCTGCTTTCGCGTTGACACAGAGAGAGGTCAAGGTCAAGGACGACGA	1118

Mon Sep 13 10:31:08 2004

us-09-939-408a-12.rnpb

Page 20

Db 1005 CCTCGAGGAGAACCGCTTTGCTGTCACCATGAGGAGGTCAAGGTCAAGGACGACGA 1064
QY 1119 GGGCAATTCTTCCGCGAGACCGGTACCCGCTCCGACCGTCCAGTTCTTCGGCCCGCT 1178
Db 1065 GGGCAATTCTTCCGCGAGACCGGTACCCGCTCCGACCGTCCAGTTCTTCGGCCCGCT 1124
QY 1179 CGTGAAGACATGATGACAGGCTTACTCGACTCTCTCGCTCAGAACACA---GAGCAG 1235
Db 1125 CGTCAAGCAGCTCATTCAGGCCACGCGGTCTCTCATCATGAGGCCGCGCATGACGAC 1184
QY 1236 CGACAAACCGCTCTCTCGAGCTGAGAGAACAAAGCAGACCGCGACCGCGCACTTCCAGGC 1295
Db 1185 CGACAAACCGCTCTCATTCAGAGCTGAGAACAAAGCTTGCACACCGCGCAATTTCCAGGC 1244
QY 1296 GTCGGCTGTCTGATTTGATGAGAGAACAGGCTCGACTCGCCTCATCGGCAAGCT 1355
Db 1245 TGCCTGTGTGCGCAACACATGAGAGAAAGCTCGGCTCGGCTCGCCAGATCGGCAAGCT 1304
QY 1356 CAATTTCACGCAGTGCACCGAGTTGCTCAACGCTGCAATGAAACCGCGCTGCTTCGTG 1415
Db 1305 CAATTTCACGCAGTGCACCGAGTGTCAACGCGGATGAAACCGCGCTGCTTCGTG 1364
QY 1416 CCTCGCTGCGAGACCGCGTCTCAATATCAAGGCAAGGCTTGGACATTCACATCGC 1475
Db 1365 CTTCCGCGCGAAGACCTCTGCTCTCTTCTTCACTGCAAGGCTCGACATCGCCGCTGC 1424
QY 1476 TGCCTACGCTTGAAGCTCGGCAACCTTGCCACCGGTCATACCTTCGTCCAGCCGC 1535
Db 1425 GCGGTACACCTGCGAGTTGGAGACCTGCGCAACCTGTGACGACGATGCCAGCGGC 1484
QY 1536 AAGATGAGGACACCGGCGCTCAACTCGCTGCTCATCTCCGCGCGCGCACTGCCGA 1595
Db 1485 TGAATGAGGACACCGGCGCTCAACTCGCTGCTCATCTCCGCTGTGCAACCGA 1544
QY 1596 GGGCAACGACGTCCTTCT 1655
Db 1545 GTCACACGACGTCCTTCT 1604
QY 1656 CGACCTCCGCGGATGAGCTCGACTTCAGAACAGATTGACCCGCTTCCGACTCT 1715
Db 1605 CGACTTGGCGGACCGAGTTGAGTTCAAGAACAGTTGGCCCAAGCTCTCTCTCTCT 1664
QY 1716 CTTCCAGACGACCTCGGCACTGGCTGAGCTCA-----ACGCACTTGGCT 1763
Db 1665 CATTCACGACCTTGGCTCCGCTATGACGCTCGAACCTGCGGACGAGCTGTGCA 1724
QY 1764 CGAGTCAAGAAAGCGCTCAACAAAGCTCTTGAGACAGACGACGTAACCTTGAGGC 1823
Db 1725 GAAGGTGAACAAGCGCTCGCAAGCGCTCGACAGACCACTCGTACGACCTCGTCC 1784
QY 1824 GGGCTGAGACGACGCTTCTCTGTAAGCGACCGGACCGGTGAGAGTCTCTCTCTCTCT 1883
Db 1785 GGGCTGAGACGACGCTTCT 1844
QY 1884 GGCCTTGGCAAGCTCACTTACTGCGGTCAACGCGTGAAGTTGCTCGGCGGAGAA 1943
Db 1845 GTC-----GCTCTGCTCGCGCGCGCTCAAGCTTGAGAGTGGCGCGCGCGAGTC 1895
QY 1944 GGGCATCTCGCTCAAGCGCGGAGTGCAGAACCGTTTGACAGACGCGCTTTCGAGGC 2003
Db 1896 GGGCATCTCGCTCAAGCGCGGAGTGCAGAACCGTTTGAGTTCGCGCGCTGACCTGTC 1955
QY 2004 GCGGCGACGACATCTCTGCGCGGACGCGGCTCTTACTGCTTGTGCGGAGGA 2063
Db 1956 GCGGCGCTCTCTGATCTCTGCGCGGACCTCATGATCTCTAGCCTGTGTCGAGGA 2015
QY 2064 GCTCGGCTGAGAGCGCGCGGCGGACGTTTGTGCGGCTGACAGAGACATCGG 2123
Db 2016 GCTTGGCTCAAGGCGCGCGGAGAGCTTCTCTCGCAAGCAAGAGTGAAGATCGG 2075
QY 2124 GAGCAAGCTTCGCGCATCTTACGAGGCAATCAAGAGCGCGGATCAACGATCTCTCT 2183
Db 2076 CTCGAAGCTTCGAAGTCTACGAGGCAATCAAGTGGGCAAGATCAACAGCTCTCTCT 2135

QY 2184 CAAGATGCTCGGCTA 2198
Db 2136 CAAGATGCTCGCTTA 2150

Search completed: September 12, 2004, 12:22:08
Job time : 1583.32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 11, 2004, 13:17:03 ; Search time 9123.4 Seconds
(without alignments)
7917.733 Million cell updates/sec

Title: US-09-939-408a-12
Sequence: 1 ctctgcctcactctaaccc.....gagttcaaaaaaaaaaaaaa 2419

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	em_estba.*
2:	em_esthum.*
3:	em_estin.*
4:	em_estmu.*
5:	em_estov.*
6:	em_estpl.*
7:	em_estro.*
8:	em_hnc.*
9:	gb_est1.*
10:	gb_est2.*
11:	gb_hnc.*
12:	gb_est3.*
13:	gb_est4.*
14:	gb_est5.*
15:	em_estfun.*
16:	em_estom.*
17:	em_gse_hum.*
18:	em_gse_inv.*
19:	em_gse_pln.*
20:	em_gse_vrt.*
21:	em_gse_fun.*
22:	em_gse_mam.*
23:	em_gse_mus.*
24:	em_gse_pro.*
25:	em_gse_rtd.*
26:	em_gse_pmg.*
27:	em_gse_vrt.*
28:	gb_gse1.*
29:	gb_gse2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	400.2	16.5	2598	11	AY103647 Zea mays
2	386.2	16.0	2299	11	AY106831 Zea mays
3	289.2	12.0	1864	11	AY104679 Zea mays
4	268	11.1	859	14	CB632830 OSJED11F

C	5	266.8	11.0	866	14	CB645202	CB645202 OSJNEB06M
C	6	266.6	11.0	860	14	CB676595	CB676595 OSJNEB120
C	7	265.6	11.0	889	14	CB648876	CB648876 OSJNEB12H
C	8	265.4	11.0	864	14	CB649712	CB649712 OSJNEB13M
C	9	265.2	11.0	867	14	CB643557	CB643557 OSJNEB04F
C	10	265.2	11.0	867	14	CB651503	CB651503 OSJNEB16H
C	11	265	11.0	860	14	CB658123	CB658123 OSJNEB130
C	12	263.8	10.9	891	14	CB651685	CB651685 OSJNEB16L
C	13	263.8	10.9	893	14	CB648470	CB648470 OSJNEB16L
C	14	263.6	10.9	868	14	CB673619	CB673619 OSJNEB08D
C	15	263	10.9	858	14	CB647704	CB647704 OSJNEB10L
C	16	262.6	10.9	857	14	CB677847	CB677847 OSJNEB15D
C	17	262.6	10.9	868	14	CB673025	CB673025 OSJNEB07F
C	18	262.2	10.8	855	14	CB649386	CB649386 OSJNEB13E
C	19	262.2	10.8	855	14	CB643494	CB643494 OSJNEB04D
C	20	262	10.8	907	14	CB649639	CB649639 OSJNEB13M
C	21	262	10.8	862	14	CB629973	CB629973 OSJNEB06K
C	22	261.8	10.8	864	14	CB629973	CB629973 OSJNEB15K
C	23	261.6	10.8	855	14	CB675774	CB675774 OSJNEB11L
C	24	261.4	10.8	851	14	CB666100	CB666100 OSJNEB12N
C	25	260.4	10.8	856	14	CB654478	CB654478 OSJNEB07A
C	26	258.2	10.7	853	14	CB644159	CB644159 OSJNEB05D
C	27	257.4	10.6	837	29	CG209714	CG209714 OGMAB587H
C	28	255.8	10.6	846	14	CB644836	CB644836 OSJNEB06E
C	29	255.4	10.6	861	14	CB650414	CB650414 OSJNEB01H
C	30	255.4	10.6	864	14	CB669223	CB669223 OSJNEB01H
C	31	255	10.5	846	14	CB641706	CB641706 OSJNEB01H
C	32	254.8	10.5	840	14	CB651026	CB651026 OSJNEB15N
C	33	254.8	10.5	845	14	CB650590	CB650590 OSJNEB15B
C	34	254.8	10.5	845	14	CB651106	CB651106 OSJNEB15P
C	35	254.8	10.5	845	14	CB651730	CB651730 OSJNEB16N
C	36	254.8	10.5	845	14	CB665372	CB665372 OSJNEB11L
C	37	254.8	10.5	928	14	CB644050	CB644050 OSJNEB05B
C	38	254.8	10.5	941	14	CB672850	CB672850 OSJNEB07B
C	39	254.6	10.5	847	14	CB676570	CB676570 OSJNEB120
C	40	254.6	10.5	862	14	CB64327	CB64327 OSJNEB08L
C	41	253.8	10.5	845	14	CB644800	CB644800 OSJNEB06C
C	42	253.8	10.5	852	14	CB666405	CB666405 OSJNEB13E
C	43	253.2	10.5	867	14	CB651444	CB651444 OSJNEB16G
C	44	253.2	10.5	845	14	CB675728	CB675728 OSJNEB11K
C	45	253.2	10.5	865	14	CB642929	CB642929 OSJNEB03F

ALIGNMENTS

RESULT 1	AY103647	2598 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	AY103647				
DEFINITION	Zea mays PCO142084 mRNA sequence.				
ACCESSION	AY103647				
VERSION	AY103647.1	GI:21206725			
KEYWORDS					
SOURCE					
ORGANISM					
	Zea mays				
	Zea mays				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
REFERENCE	1 (bases 1 to 2598)				
AUTHORS	Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanefey,M., Morgante,M. and Tingey,S.V.				
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2 (bases 1 to 2598)				
AUTHORS	Coe,E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				

maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES

source

1..2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="maizedb:638378"
/db_xref="taxon:4577"
/clone_id="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

ORIGIN

Query Match 16.5%; Score 400.2; DB 11; Length 2598;
Best Local Similarity 53.4%; Pred. No. 6,3e-45;
Matches 1095; Conservative 0; Mismatches 903; Indels 53; Gaps 10;

184 CAGCTCGAGTCGTGACGAGCTCTCAGCGACCCCGACGAGCGTGTGAGCTCAGC 243
Db CACTGACGAGGTGAAGCCCATGTGGGCGAGGCCCGGACCTGTGTAAGTCAG 244
QY GGGTACAGCTCAACCTCCG-----TGACGTTGTGGGCGCCCGGACAGGGGCGCAGG 297
Db GGTTCACCTCTCGCTCGCCGACGAGTGGCGCCGCTGCTCCGACAGACGCGTCGCG 304
QY GTCCGCGTCCAGAACGACGACGAGATCGGACGCGTGCAGAAAGGCTGACTCTC 357
Db GTGCGCGTCAAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 364
QY AAGGCCCAAGCTTCAAGACTCG-----TTTACGAGATCACCAAGGCTTTCGATGCTCG 411
Db CTGACTGATCGCCACGCGCGGACATCTACGCGCTCACACGCGCTTCGGGCGCAC 424
QY GCCGACGAGACTGAGATGATGACGCTCCGAGAGGCGCTCATGAGCACCACTC 471
Db TCCACGCGCGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 484
QY TGGGCGGTAGACCGGACGTCGCTCTTCAAGCTTCGAGACGCGGCTGAGAACAG 531
Db GCGCGA-----ATCTTCGACACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 520
QY CTTCGCGCTGAGATGCTCCGCGCGCATGATCGCGTCAACTGCTCAAGCTGCG 591
Db CTGCGCTGAGAGTCAACCGCGCGCATGCTGAGCGCATCAACCTCTCTCAAGGCG 580
QY CACTCGGCGCTGCTGCTCTTAGGCGCTCAACCACTTCTGAACCAACCGCATC 651
Db TACTCGGAGATCGCTTGAAGTCTTCAAGCGCATCAAGAGCTGCTCAACCGGCTC 640
QY ACCGCTATGTCCTCCCTCCGCGGCTCATCTCGAGTGGGAGACCTCAAGCGCTCG 711
Db AACCCCTGCTGCTCGCTCGGCGGACATCAACCGCTGGGCGAGCTGCTCGCTCG 700
QY TACATCGCGGCGCATACCGGTCACCGGACGTCAGAGTTACGTTTTCAGAGGA 711
Db TACATCGCGGCGCTATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGG 760
QY ACCGAAAGATCATTTTGGCGGAGGCGATCTGCTTGTGTCAGAGGAGTCTC 831
Db AAGGTGAGACCGCGCGAGGCGGCTTCAAG---ATCGCGGATCAAGGCGGCTTCTCAAG 817
QY CTCGCGCGGAGAGAGGCTGCTGCTGCTCAACGAGAGCGCGCTTCGCTGAGAGG 891
Db CTCGCGCGGAGAGAGGCTGCTGCTGCTCAACGAGAGCGCTTCGCTGAGAGG 877
QY ACCCTCACTGACAGCTGACATGCTCTGCTCTGAGAGGCTTGAAGGCTCTC 951
Db

878 GCCACGGATGATGACAGCCCAACGCTGCGGCTGCTGAGAGCTCTCGCGCTC 937
QY ACCGTGAGAGCATGATGCTCGCCAGAGAGGCTGTTCCGCGCTTATCAAGAGTCTGC 1011
Db TTGTGAGAGTATGAACGCGAAGGCCAGATCAAGGACCACTGACCCCAACTGTAAG 997
QY CGCCCGCACCGCGGCGAGGTCGAGTCGCGGCGCAATCCGACGCTCTTCCGCGCG 1071
Db CAC---CACCGGGGCTCATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1054
QY TCGTTTCCGTTGAGCAGAGGAGGATCAAGTCAAGAGCAAGAGGAGGATTTCTTGC 1131
Db TCCTT---CATGAGAGGCGCAAGAGTTGAACGAGCTGAGACCGGCTGTAAGCCAG 1111
QY CAGACCGCTACCGGCTGCGACGTCGAGTCAAGTTCCTGCGCGGCTCTGAGAGCATG 1191
Db CAGACAGTATGCGGCTCGGACGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1171
QY ATGACGCTTACTGATCTCTCGCTGAGAAACAAGACGACCGGACCAACCGGCTCTC 1251
Db CGGCGCGCACCAAGTTCATGAGGCGAGGTCGAAGTCCGTAACGACCAACCGGCTCATC 1231
QY GACGTCGAGAAACAAGACAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1311
Db GACGTCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291
QY TCGATGAGAAAGACAGGCTGCACTGCGCTCATCGGCAAGCTCACTCAAGAGTGC 1371
Db TCCATGAGAAAGCGCGCGCTCGCATGCGCAACTCGGCAAGCTCATGTTGGGCAATC 1351
QY ACCGATGCTCAAGCTGCTCATGAACCGCGGCTGCTTGTGCTGCTGCTG---CCGAG 1428
Db TCGAGCTGCTCAAGAGTTCATCAACAAGCGGCTCACTTCAACCTGCGCGGCGGCGG 1411
QY GACCGTGCCTCAACTATCAAGGCAAGGCTTGAACATTCACTGCTGCTTACGCTTGC 1488
Db AACCCAGCTGAGACTAGGCTTCAAGGCAAGGCTTCAAGGCTGCTTCACTGCTTGC 1471
QY GAGCTGCGCACCTTCCGACCGGCTCATCTTGTTCAGCGCGGAGAGATGGCAAC 1548
Db GAGCTCCAGTACTGCGGCAACCCCATCAACAACGCTGAGAGCGGCGGAGCAAC 1531
QY CAGGCGCTCACTGCTGCTCATCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1608
Db CAGAGCTGAATCTCTCGGCGGCTGCTTCCGCGGAGAGAGAGCGGCGGAGTGCATC 1591
QY CTTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1668
Db CTGAAGCTCATGCTGCTCACTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1651
QY ATGAGGCTCGACTTCAAGAGAGAGTTCAGACCGCTTCTCCGACTCTCTCAAGAGAG 1728
Db CTGAGAGAGATCAATCAAGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
QY CTGCGCACTGCGCTCGAGCTCAA---CGCACTTGCCTGAGAGTCAAGAGAGAGAG 1785
Db CTGACATGAACCCCTCGGCGGAGCTCTCCAGGCGCGCTTCAAGCAAGAGAGAGCTGATC 1771
QY AAGGCTTCAG 1845
Db AGCGCATCAACCGGAGGCGGCTTCAAGTACGTAAGAGAGAGAGAGAGAGAGAGAG 1831
QY TACCGCA---CGGACAGCTGCTGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1904
Db CCGTGAATGAG 1891
QY TACTGCGCTCAAGCGTGAAGGTTCTCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1964
Db GAGGAG 1951
QY GGTGCGCAACCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2024
Db CTGCGCAAG 2010

QY 2025 GCCGCGCAGCGCGCTCCTGTACTGCTTGTGCGGAGAGAGCTCGGCTGCAGAGCGCGCCG 2084
DB 2011 GAACCGGATCGCGGACACCGCGCTGCTCCCGCTTACCGCTTCTGCGCGAGAGAGCTCGG 2070
QY 2085 CGGCGACGCGTTTGTGCGGCGTGCAGCAGAGAGACGATCGGAGCAACGCTTCGCGCATCTA 2144
DB 2071 CTGCGTGTCTTCCAGCGCGAGAGGCTCAAGTCCCCCGCGAGAGATGCAACAAGTGTCT 2130
QY 2145 CGAGCGCATCAAGAGAGCGCGCGCATCAACACAGCTCTCGTCAAGTGTCTCGGCTAAGGCC 2204
DB 2131 CGTCGCGCATCAACCGAGGCGAGAGCTGTGAGATCCCATGCTGTGAGTCTCAAGAGATGGA 2190
QY 2205 GAGCAAGCTCT 2215
DB 2191 CGGCAAGCGCG 2201

RESULT 2
AY106831 2299 bp mRNA linear HTC 16-OCT-2002
LOCUS Zea mays PC0142079 mRNA sequence.
DEFINITION AY106831
ACCESSION AY106831 GI:21209909
VERSION HTG.
KEYWORDS Zea mays
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2299)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitelc,M.S.,
Arthur,L.W., Hanley,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2299)
Coe,E.H.

JOURNAL Direct Submission
REFERENCE Submitted (25-APR-2002) Maize Mapping Project, University of
AUTHORS Misouri, Columbia, MO 65211, USA
TITLE If you are interested in getting corresponding physical clones,
JOURNAL these are publicly available from ZmDB and may be found by BLAST
searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
COMMENT www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Malbot, Stanford or Pat
Schabale, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES
Location/Qualifiers
1..2299
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

ORIGIN
Query Match 16.0%; Score 386.2; DB 11; Length 2299;
Best Local Similarity 55.1%; Pred. No. 5e-43;
Matches 950; Conservative 0; Mismatches 753; Indels 22; Gaps 9;

QY 503 TCAGCTTCGAGCGCGCTTCGAGAAACAGCTTCGCTGAGGTCTCCGCGCGCCATCG 562
DB 246 TCTTCGCGAAGCGCTTCGAGCGCGCAACAGCTTCGCTTCGAGGTTCGCGCGCGCAATGC 305
QY 563 TCATCGCGTCAACTGCTCAGCGGTGCGCACTCGCGCGCTCGCTCTTGAGG 622

DB 306 TGTTCGATCAACACCTTCTCTCAGGAGCTACTTCGCGATTCGCTTCGATCTCTGAGG 365
QY 623 CGCTACCACTTCTTGAACACCGCATCAAGCCATGCTCCCTTCGCGCTCACTT 662
DB 366 CCATCAACCACTCATCAACACCGCGCTCA--GCCGTCTCGCGCTCGCGGAGCATCA 424
QY 663 CGGCGTGGGAGACTTACGCGCTCTCTGATACATCGCGCGCGCATCAAGGTCAACCCG 742
DB 425 CGGCTTCGCGAGCTCTGCTCCGCTGTCTACATCGCGCGCTCATACGCGCGCTCA 484
QY 743 AGTCAAGTTCAGGTTTTCGACGAGGAAACGAGATCATGTTTCGCGCGAGCCA 802
DB 485 ACGCGAAGCGCGTCAAGGTTGACGAGAGAGAGAGAGAGCGCGCGAGGCTTCAAGCGG 544
QY 803 TCTCGCTTTTGTCTGAGAGAGAGTCTCTCGCGCGCGAGAGAGAGTCTGCTGTA 862
DB 545 CGGCGATGAGAGG---CGGCTTTCAGGCTCAACCCAGAGAGGCGCTGCGCATGCTCA 601
QY 863 ACGGAGCGCGCTCCGCTCGATGCGAGACCTCACTGACAGATCGCATGCTCT 922
DB 602 ACGGAGCTCCGTGGGCTCGCGCTCGCGCATGCTGCTTGAACGCGCATGCTCTG 661
QY 923 CGCTCTCTCGAGCGCTTTCAGCGCTCTTCAGGTGAGGCGCATGCTGCGCATGAGGCT 982
DB 662 CGCTCTCTCGAGGCTCTGCTCCGCTCTCTCTGCGAGGATGATGAGCGAAGCGCGAGT 721
QY 983 CGTTGCGCGCTTCATCAGACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1042
DB 722 ACACGAGCACTTACCTCAAGCTCAAGAC--CACCGGCGCTCATGAGCGCGCG 778
QY 1043 GCAATCGCGAGCGCTCTCTTCGCGCTGCTGTTGCGGTTGACAGAGAGAGGTCA 1102
DB 779 CCATATGAGAGACATCTGATGAGAGCTCTT---CATGAGAGAGCGCGAGAGGTGA 835
QY 1103 AGTCAAGAGAGAGAGAGGAGATTTCTTCCGAGAGCGCGCTTCGCGCGCGCGCTC 1162
DB 836 ACGCATGAGCGCGCTGCTGTAAGCGAGAGAGAGAGAGAGAGCTCGAGCTGCGCGC 895
QY 1163 AGTTCCTCGCGCGCTCTGAGAGAGATGACAGCGCTACTCTCTCTCTCTCTGAGA 1222
DB 896 AGTGCCTCGCGCGCGCGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
QY 1223 ACAACAG 1282
DB 956 TCAATCGCTCAACAG 1015
QY 1283 GGAATTCAGAGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1342
DB 1016 GCAATTCAG 1075
QY 1343 TCATCGAG 1402
DB 1076 ACATCGGAG 1135
QY 1403 GCTGCTCTCTGCT 1459
DB 1136 GCTGCT 1195
QY 1460 TGAATTCATCAATCGCT 1519
DB 1196 CGAGATCGCGAG 1255
QY 1520 CTTGCTTCAGCGCGCGAG 1579
DB 1256 ACCAGCTTCAG 1315
QY 1580 CGGCGCGAG 1639
DB 1316 CAG 1375
QY 1640 GCAAGCTTCAG 1699

Db 1376 CGCTGTGCGACGCGGTGACCTGTGGACACTGTGAGGAAAGCTCAAGAGCCCGCTCAAGA 1435

Qy 1700 CGCTTCCCGGACTCTCTCTACAGACACTCGGCACTGCGCTTCGACGTCAAGAC--- 1756

Db 1436 GCTGCGTATGCGGTGGCCAGAGGTGTGTACCAACGAGCTTCGCGCGGCACTTCACA 1495

Qy 1757 TTGGCTGAGGTCAAGAGCGGCTCAAGAGGCTTCGAGAGAGAGAGAGAG--- 1813

Db 1496 GCGGCGCTTCAGAGAGAGAGCGCTGTGTACCGCCATTCAGCGGAGGCGGTGACGGGT 1555

Qy 1814 ACCCTGAGCGCGCTGAGAGAGAGCTTCTGTGACCGGACCGGACCGCTGTGAGCTTC 1873

Db 1556 ACTACGACGAGCGCTGTGAGAGCGCACTGCGCCCTGTATGAGAGATCCGGGCGGTGTGG 1615

Qy 1874 TCTGTCTCTGCGCTTCGCAAGCTACCTTACTGCGCTTCAGAGCGGTGAGAGGTGCT 1933

Db 1616 TGAACCAAGCGCTTCGCAAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675

Qy 1934 CGGCGGAGAGAGCGCTATCTGTGACAGCGGAGAGTGC-GCACCGCTTGTGAGAGAGCGCG 1992

Db 1676 AGATCAACAGAGTTCAGAGAGAGAGTGCAGAGAGCGCTGCGGAGAGATGAGAGCGCGCC 1735

Qy 1993 TCTTTCGAGCGCGCGCGGAGAGAGCTATCTGTGAGCG---GGGACGCGCGCTGTACTGT 2050

Db 1736 GCTGTGCTTCGAGAGCG 1795

Qy 2051 TGTGTGCGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110

Db 1796 ACCGCGGTGACCGCTTCATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855

Qy 2111 AGAGAGAGATGCGGAG 2170

Db 1856 TCAAGTCCCGCGGAG 1915

Qy 2171 ACCAGCTCTCTGCTGAG 2215

Db 1916 TGAACCCCATGCTGTGAG 1960

RESULT 3

AY104679 1864 bp mRNA linear HTC 16-OCT-2002

LOCUS Zea mays PC0142078 mRNA sequence.

DEFINITION AY104679

ACCESSION AY104679.1 GI:21207757

VERSION

KEYWORDS

SOURCE

ORGANISM

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 1864)

Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, J.W., Hainey, M., Morgante, M. and Tingey, S.V.

Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes

Unpublished (2002)

2 (bases 1 to 1864)

Coe, P.H.

Direct Submission

Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

Journal

COMMENT

If you are interested in getting corresponding physical clones, these are publicly available from ZmB and may be found by BLAST searching at MSL, maizemap.org; ZmB, www.zmbl.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmB: www.zmbl.iastate.edu.

Location/Qualifiers

1..1864

/organism="Zea mays"

/mol_type="mRNA"

ORIGIN

Query Match 12.0%; Score 289.2; DB 11; Length 1864;

Best Local Similarity 57.8%; Pred. No. 7.8e-30;

Matches 594; Conservative 0; Mismatches 423; Indels 11; Gaps 4;

1376 CGCTGTGCGACGCGGTGACCTGTGGACACTGTGAGGAAAGCTCAAGAGCCCGCTCAAGA 1435

1700 CGCTTCCCGGACTCTCTCTACAGACACTCGGCACTGCGCTTCGACGTCAAGAC--- 1756

1436 GCTGCGTATGCGGTGGCCAGAGGTGTGTACCAACGAGCTTCGCGCGGCACTTCACA 1495

1757 TTGGCTGAGGTCAAGAGCGGCTCAAGAGGCTTCGAGAGAGAGAGAGAG--- 1813

1496 GCGGCGCTTCAGAGAGAGAGCGCTGTGTACCGCCATTCAGCGGAGGCGGTGACGGGT 1555

1814 ACCCTGAGCGCGCTGAGAGAGAGCTTCTGTGACCGGACCGGACCGCTGTGAGCTTC 1873

1556 ACTACGACGAGCGCTGTGAGAGCGCACTGCGCCCTGTATGAGAGATCCGGGCGGTGTGG 1615

1874 TCTGTCTCTGCGCTTCGCAAGCTACCTTACTGCGCTTCAGAGCGGTGAGAGGTGCT 1933

1616 TGAACCAAGCGCTTCGCAAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675

1934 CGGCGGAGAGAGCGCTATCTGTGACAGCGGAGAGTGC-GCACCGCTTGTGAGAGAGCGCG 1992

1676 AGATCAACAGAGTTCAGAGAGAGAGTGCAGAGAGCGCTGCGGAGAGATGAGAGCGCGCC 1735

1993 TCTTTCGAGCGCGCGCGGAGAGAGCTATCTGTGAGCG---GGGACGCGCGCTGTACTGT 2050

1736 GCTGTGCTTCGAGAGCG 1795

2051 TGTGTGCGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110

1796 ACCGCGGTGACCGCTTCATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855

2111 AGAGAGAGATGCGGAG 2170

1856 TCAAGTCCCGCGGAG 1915

2171 ACCAGCTCTCTGCTGAG 2215

1916 TGAACCCCATGCTGTGAG 1960

1022 CGGCGAGGTGAGAGTGTGAG 1081

355 CGGAGAGATGAGAGCGCGCGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414

1082 TTGAGAGAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141

415 AGCTGGCCMAAGCTGTGAG 472

1142 ACCGCTCGGAGAGTGTGAG 1201

473 AGCGCTCGGAGAGTGTGAG 532

1202 ACTGAGCTCTGCTGTGAG 1261

533 CCAAGTCATGAG 592

1262 ACAAG 1321

593 GTGGCAG 652

1322 AGACAG 1381

653 ACACCGCTGTGAG 712

1382 TCAAGCGTGTGAG 1438

713 TGAAGAGATGAG 772

1439 TCAAGTGTGAG 1498

773 TGAAG 832

1499 ACCTTGCAG 1558

/db_xref="MaizeDB:638373"

/db_xref="taxon:4577"

/clone_lib="Maize Mapping Project/Dupont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

Db 833 TCCTGGCAACCCGGTGAACAACACGCTCCAGAGCGCGAGACAGACAACAGAGCTGA 892
 QY 1559 ACTCGCTCGCTCTCATCTCCCGCGCGCCGACATGCGCGAGAGCCCAAGACGTCTCTCTCC 1618
 Db 893 ACTCCCTGGGCTCATCTCTGTCAGAAAGACCGCGAGGACGTGACATCTTGAAGCTGA 952
 QY 1619 TTCTGCTCGACCTGTACTGACAGCTCCAGCGCCCTGACCTCCGCGCATGAGCTCG 1678
 Db 953 TGACGTCCACGTTCTGATCGCGCTGTGTGCGAGGCATCGACCTGGGCACTTCAGAGAGA 1012
 QY 1679 ACTTCAAG 1686
 Db 1013 ACGTGAAG 1020

RESULT 4
 CB632830
 LOCUS CB632830 859 bp mRNA linear EST 08-APR-2003
 DEFINITION OSIIEB1F05.f OSIIEB Oryza sativa (indica cultivar-group) cDNA
 clone OSIIEB1F05 5', mRNA sequence.

ACCESSION CB632830.1 GI:29627819
 VERSION CB632830
 KEYWORDS EST.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 859)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea

UNPUBLISHED (2003)

CONTACT: Rod Wing

UNIVERSITY OF ARIZONA
 ARIZONA GENOMICS INSTITUTE

BIOLOGICAL SCIENCES WEST, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: http://genome.arizona.edu

PCR PRIMERS

FORWARD: gta aaa cga cgg cca gtc

BACKWARD: gga aac agc tat gac cat g

Place: 11 row: P column: 05

Seq primer: gta aaa cga cgg cca gtcg.

Location/Qualifiers

1..859

/organism="Oryza sativa (indica cultivar-group)"

/mol_type="mRNA"

/cultivar="IR36"

/db_xref="taxon:39946"

/clone="OSIIEB1F05"

/tissue_type="leaf"

/dev_stage="3 week"

/lab_host="DH10B"

/clone_1b="OSIIEB"

/notes="vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (P06-6-3)"

Db 61 GAGATCTCTGAGGCCATCAACCAAGCTGTCTCAACACCGGCGTCAAGCCGCTGCTCCGCTC 120
 QY 670 CGGCGCTCATCTGGGCGCTGGGCGGACCTGAGCCGCTCTGTACATGCGCGGCGCATC 729
 Db 121 CATTGACATCAACCGGCTCCGCTGACCTGCTTCCCTGTCTCATATGCGCGCTCATC 180
 QY 730 ACCGCTACCCCGACGTCAGAGTTCAAGTTTGTGACAGAGGAAACGAGAAATATATTTT 789
 Db 181 ACCGCGCCCGCCCAAGCCAGAG---CCATCTGCGCCGACAGGAGGAGTGAAGCGCC 237
 QY 790 GCGCGGAGGCGCATCTGCTCTTTGTGTCTGAGGAGCGCTCCGCGCCGAGAGGAGT 849
 Db 238 GAGGCGTTCAAGCTGCGCGGATGAGAGGTGCTTTTCAAGCTGAACCCCAAGAGT 297
 QY 850 CTGCTGTGTCAACGAGACGCGCTCTCCGCTGATGAGGACCTTCACTCTGACGAC 909
 Db 298 CTGCGCATCTGTAATGACAGCTCGGTGGGCGCGGCTCGCGGACCGATGATGAC 357
 QY 910 TTGCACATGCTCTGCTGCTCTGCGAGGCTTGAAGGCTCTCAAGGTGAGAGGCGATGCTC 969
 Db 358 GCGACATCTGCGCGCTCTGCTGCGAGGCTCTGCGGCGGTCTTCTGCGAGGTGATGAC 417
 QY 970 GCGCAGAGGCTCTGCTGCGCGCTTCAATCCAGCAGCTTCCCGCCGACCCCGCGCAG 1029
 Db 418 GCGCAGCGGAGTACACGACACCTGACCCAGAGCTGAAGAC---CACCTGGGTG 474
 QY 1030 GTGAGAGTGTGCGCGGACATCCGACGCTCTTTCCGCTGCTGTTGCGGTGAGCAG 1089
 Db 475 ATCGAGGCGCGCGCATATGAGAGACATCTTCCCGGAGACTGTT---CATAGCCAC 531
 QY 1090 GAGGAGAGGTCAAGTCAAGAGACGAGAGGAGCTTTCTGCGCAGAGCGCTACCCGCTC 1149
 Db 532 GCGAAGAGTGAAGATGAGACCGCTGCTGAGACCGAAGAGAGAGTACGCGCTC 591
 QY 1150 GCGAGTGTGCTCAAGTCTCTGCGCGCTCTGTGAGACATGATGACGCTTCAACT 1209
 Db 592 CCGAGTGTGCGCGCATGCTGCTGCGCGCGGACGATGAGGTCATCCGCGCGCACCAAGTCC 651
 QY 1210 CTCTGCTGAGAAACAAGACGACGACGACGACGACGACGACGACGACGACGACGACG 1269
 Db 652 ATCGAGCGGAGTGAATCTGCTGAGACGACGACGACGACGACGACGACGACGACG 711
 QY 1270 ACCGCGCAGCGCGGCACTTCCAGCGCTGCTGCTGATTTGATGAGAGACGACG 1329
 Db 712 GCGCTCAACGCGGCACTTCCAGCGGACCCCATCGGTGTGCTCATGAGCAAGCGCGT 771
 QY 1330 CTGCGACTGCGCTTATGCGGACGTCATCTTACGACGACGACGACGACGACGACG 1389
 Db 772 CTGCGCATGCGCAATGCGGAGCTCATGTCGCGGAGTCTTCCGAGCTCTGAGCGAA 831

QY 1390 GCCATGAACCGCGGCTGCTTGTGCGC 1417
 Db 832 AGCTTACACACGCGCTGACCTCCAC 859

RESULT 5
 CB645202/c 866 bp mRNA linear EST 08-APR-2003
 LOCUS OSJNB06M01.r OSJNB Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNB06M01 3', mRNA sequence.
 DEFINITION CB645202.1 GI:29640193
 VERSION CB645202
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 866)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea

JOURNAL COMMENT

Unpublished (2003)
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 06 row: M column: 01
 Seq primer: gga aac agc tat gac cat g.

FEATURES

source

1. .866
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSUNB06M01"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSUNB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"

ORIGIN

Query Match 11.0%; Score 266.8; DB 14; Length 866;

Best Local Similarity 59.7%; Pred. No. 8.7e-27;

Matches 503; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

QY 832 CTGCGCCCGAAGAGGGTCTCGTCTGTCACGACGCGCGTCTCGCCTCGATGCGG 891
 DB 865 CTGAACCCCAAGAGAGTCTCGCATCTGCATGACGACGTCGCGGGTGGCGCTCGCG 806
 QY 892 ACCCTGAGTGTGACGACGTGACATGCTGCTGCTGCTGCGACGCGCTTACGCGCTCTC 951
 DB 805 GCCACCTGATGTTCACGCGCAACATCTCTGCGCTGCTGCGAGGTCCTCTCGCGGATG 746
 QY 952 ACCTGAGAGCCATGCTGCGCCAGACGAGGCTGTTGCGCGCTTATCCAGAGCTGTC 1011
 DB 745 TTCTGAGAGGTATGACGCAAGCGGAGTACACGACCACTGACCCCAACAGCTGAG 686
 QY 1012 CGCCCGACCCCGCGGACGTCGAGTTCGCGGCAACATCCGACGCTCTTTCCGCGTCG 1071
 DB 685 CAC--CACCCTGGGTGATGACGAGCGCGCGCATCATGAGACATCTGCGCGGAGC 629
 QY 1072 TCCTTGCCTTACGACGAGAGAGAGTCAAGTCAAGACGACGAGGCAATCTTCGCG 1131
 DB 628 TCCTT---CATGAGCAACCCCAAGAGTGAAGAGATGAGACCGCGTCTGAACCGAAG 572
 QY 1132 CAGGACCGCTACCCCTCCGACAGTGGCTCAATGTTCTCTGCGCGCGCTCTGAGAGACATG 1191
 DB 571 CAGGACAGGTACCGCTCTCCGACAGTCCGCGAGTGGCTCGCGCCGCAATGAGATGATC 512
 QY 1192 ATGACGCGCTACCTCGACTCTCTGCTCGAGAACACAGACGACGACCAACCGCTCTC 1251
 DB 511 CGGCGCGCACCAAGTTCATCGAGCGGCAAGTCAACTCGTGAAGACACACCGCATATC 452
 QY 1252 GACGTGAGAACCAAGACGACGCGCAACCGCGCAACTTCCAGGCGGTGGCTGTTCGATT 1311
 DB 451 GACGTCCACCGCGGCAAGGCGCTCCACGCGGCACTTCCAGGCAACCGCATGATG 392
 QY 1312 TCAGTGAAGAACCAAGTTCGACTCGCGCTCACTGCGCAAGTCAACTTCAAGCAATGCG 1371
 DB 391 TCATGAGCAAGCGCGCTTCTGCGCATCGCAACATCGGCAAGCTCTATGTTGCGCAATC 332
 QY 1372 ACCGAGTGTCTCAACGCTTCCATGACCGCGCGCTGCTGTGCTGCTGCTG---CCGAG 1428
 DB 331 TCCGAGCTCGTGAACGAGTTCACAAACAGGCGTGAACCTCCACCTGCGCGGAGCGCG 272

QY 1429 GACCCGTGCTCAACTATCAGGCAAGGCTTGACATTCACATCGTGTACGCTTCG 1488
 DB 271 AACCCGAGTGTGATTAAGGATTCAGGCAACGAGATCGGACGCTCTCTACTGCTCT 212
 QY 1489 GAGCTGGGCACTTGGCCACCCCGGTCACTACCTTCTCGACCGCGGAGATGGGCAAC 1548
 DB 211 GAGCTTCAGTACTCTGCGCACCCCATCAACCAATGTCCAGAGCGCGGAGCAAC 152
 QY 1549 CAGCGGTCACTGCTGCTCTCATCTCCGCGCGCGCACTGGCGGCAAGCAAGCTG 1608
 DB 151 CAGGACGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 92
 QY 1609 CTTTCTCTCTTCTGCGCTGCGACCTGTACTGACGCTCGAGCGGCTGACCTTCGCGCG 1668
 DB 91 CTCAGCTCATGACCTCCACCTACATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32
 QY 1669 ATGAGG 1674
 DB 31 CTCGAG 26

RESULT 6
 CB676595/c 860 bp mRNA linear EST 09-APR-2003
 LOCUS OSUNB012020.1 OSUNB Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION Clone OSUNB012020.1, mRNA sequence.
 ACCESSION CB676595
 VERSION CB676595.1 GI:29680320
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 860)
 Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe oryzae
 Unpublished (2003)
 CONTACT: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 12 row: O column: 20
 Seq primer: gga aac agc tat gac cat g.

FEATURES

source

1. .860
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSUNB012020"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSUNB"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

ORIGIN

Query Match 11.0%; Score 266.6; DB 14; Length 860;
 Best Local Similarity 59.8%; Pred. No. 9.2e-27;
 Matches 503; Conservative 0; Mismatches 329; Indels 9; Gaps 3;

```

QY      837 CCCGAGAGAGGCTCTCGGCTGTGCTCAACGAAACGCGCGTCTCCGCTCGATGGCGACCT 896
      |||||
DB      860 CCCCAAGAGAGGCTCTCGCATGTGCAATGGACACGCTCGTGGGTGCGCGCTCGCGGCGAC 801
QY      897 CAGTGTGACGATCGTGCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 956
      |||||
DB      800 CGTGAATGTTCAGCGCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 741
QY      957 GGAGGCGATGCTCGGCGCAAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1016
      |||||
DB      740 CAGAGGTGATGAACCGGCAACCGGATTAACCGACCACTGACCAACAAGCTGAAGAC-- 683
QY      1017 GCAACCCCGGCGAGGTGAGGTGCGCGCAACATCCGACGCTCTCTCTCTCTCTCTCTCT 1076
      |||||
DB      682 -CACCTTGGTGTGATGAGGCGCGCCATATGAGACATCTCTCTCTCTCTCTCTCTCTCT 624
QY      1077 TGGCGTTGACGACGAGGAGGAGGTCAAGGTCAAGGACAGAGGCAATTTCTCCACAGA 1136
      |||||
DB      623 ---CATGAGCCAGCCGMAAGGTGAAGATGACCCCGCTGCTGAAGCCGAGACAGA 567
QY      1137 CGGCTACCCGCTCGCAGCTGCGCTCAAGTTCCTGCGCCGCTCTCTGAGAGCATGATGCA 1196
      |||||
DB      566 CAGGTACGCGCTCCGACGCTCCGCGAGTGTGCTCGCCCGAGATCGAGTCACTCCGCG 507
QY      1197 CGGCTACCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1256
      |||||
DB      506 CGCACCAAGTTCATGAGCGCGAGGTCACTTCGTGAACGACCAACCGGATGATGACGT 447
QY      1257 CGAGAACAGACAGACCGCGCAGCGCGCACTTCCAGGCGTGTGTCTCTCTCTCTCTCTCT 1316
      |||||
DB      446 CCACCGCGGCAAGGCGCTCCACGCGGCACTTCCAGGCGACCCCATCGTGTGTCAT 387
QY      1317 GGAAGAACAGAGGCTGCACTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1376
      |||||
DB      386 GGAACAAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 327
QY      1377 GTTGCTCAACGCGGCGCATGACCGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1433
      |||||
DB      326 GCTCTGAAAGAGATTTCACACAGGCTTACCTCTCAACCTGCGCGGCAACCGCAACC 267
QY      1434 GTTGCTCAACATTCACAGGAGGCTTGAACATTCACATCTCTCTCTCTCTCTCTCTCT 1493
      |||||
DB      266 GAGCTTGAGACTAGGGTTCAAGGCGCACGAGATCCCATGCGCTCTCTCTCTCTCT 207
QY      1494 CGGCGCCTTTGCGCAACCGCGGTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1553
      |||||
DB      206 CCGATCTCTGCGCAACCGCATACCAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147
QY      1554 CGTCAACTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1613
      |||||
DB      146 CGTGAATCGCTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 87
QY      1614 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1673
      |||||
DB      86 GTCACGACCTCCACCTACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 27
QY      1674 G 1674
      |||||
DB      26 G 26

```

```

REFERENCE
AUTHORS
1 (bases 1 to 889)
Jantaasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL
Unpublished (2003)
COMMENT
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 12 row: H column: 22
Seq primer: gga aac agc tat gac cat g.
FEATURES
source
1..889
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB12H22"
/issue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_1fb="OSJNB"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"
ORIGIN
Query Match 11.0%; Score 265.6; DB 14; Length 889;
Best Local Similarity 59.8%; Pred.No.1.3e-26;
Matches 502; Conservative 0; Mismatches 329; Indels 9; Gaps 3;
QY      838 CCGAGAGAGGCTCTGGTCTGTGTCACGGAAGGCGCTCCGCTCGATGGCGACCTC 897
      |||||
DB      859 CCCAAGAGAGGCTCTGCATCTCATAGCAAGTCGTCGAGGTGCGCTCGCGCCAC 800
QY      898 AGTTCAGACGATCTCCACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 957
      |||||
DB      799 GTGATTTGAGCCCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 740
QY      958 GAGGCCATGTGCGCAGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1017
      |||||
DB      739 GAGGTGATTAAGGCAAGCGCGAGTACACGACCACTGACCCCAACACTGAAGAC-- 683
QY      1018 CAGCCCGGCGAGGTGCGCGGCAACATCCGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1077
      |||||
DB      682 CAGCTGGGTGATGAGGCGCGCGCCCATCATGAGCAATCTCTCTCTCTCTCTCTCTCT 624
QY      1078 GCCGTGAGCAGAGAGAGGTCAAGTCAAGAGCAAGGCGATTTCTTCCAGGAC 1137
      |||||
DB      623 --CATGAGCCAGCGCAAGAGGTGAACGATGAGACCCGCTGTGAAACCCGAAGAGAC 566
QY      1138 CGCTACCGGCTCGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1197
      |||||
DB      565 AGGTACGCGCTCTCGACGTCGCGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 506
QY      1198 GCTTACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1257
      |||||
DB      505 GCACCAATCTCATGAGGCGGAGGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 446
QY      1258 GAGAACAGACAGCGCGGCGGCACTTCCAGGCGTGTCTCTCTCTCTCTCTCTCTCTCT 1317
      |||||
DB      445 CACCGGCGGAGAGGCGCTCCACGCGGCACTTCCAGGCGACCCCATGCGTGTGTCATG 386
QY      1318 GAAAGAACGAGGCTGCACTCGCGCTCATGCGCAAGTCACTTCAACGACGACGAG 1377

```

RESULT 7
 CB64876/c
 LOCUS
 DEFINITION OSJNB12H22, r OSJNB Oryza sativa (japonica cultivar-group) cDNA
 clone OSJNB12H22 3', mRNA sequence.
 ACCESSION
 CB64876
 CB64876.1 GI:29643869
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

RESULT	8
CB649712/c	
LOCUS	
DEFINITION	864 bp mRNA linear EST 08-APR-2003
OSJNB1M21	r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
clone OSJNB1M21.3,	mRNA sequence.
CB648712	
ACCESSION	
VERSION	
CB649712.1	GI:29644705

Oryza sativa (japonica cultivar-group)

Ehrhartoideae; Oryzeae; Oryza.
(bases 1 to 864)

Large-scale identification of ESTs involved in the interaction

Arizona Genomics Institute
Contact: Rod Wing

Location/Qualifiers
1. .864

```

/lab_host="DH10B"
/clone_lib="OSJNEb"

```

11.0%; Score 265.4; DB 14; Length 864;

QY	835	GGCCCCGAGAGAGAGGCTCGGTCTGATGTCAACGAAACGGCCGTCTCCGCTCGATGAGCAC	894
Db	862	GACCCCAAGGAAGGCTCTGGCATGTCAATGGACACGTCTGTGGGATGGCCCTCGCGCC	803
QY	895	CTCACTTCGACGACTCGACATGCTCTGCTCTCTTCCAGGCTTGAACGGCTCTGACG	954
Db	802	ACCGTAGATGTTGCAAGCCCAATCCTTCGCGCTCTGTTCCAGAGGTCTCTCGACGATGTC	743
QY	955	GTGAGAGCGCATGATGGGCAAGAGAGGAGTCAAGTCAAGACACAGAGGCAATTTTCCGAC	1014
Db	742	TCCGAGGTGATGAAACGGACACCGAAGTACACGACCACTGACCCCAAGCTGAAGCAC	683
QY	1015	CCGCAACCCCGGCAAGTGGAGTGGCGCGCAACATCCGACGCTCTTTCGCGCTCTCG	1074
Db	682	---CACTTGGGTGATGAGAGCGCGCCGACCATATGAGACATCTTCGCGCGGAGCTCG	626
QY	1075	TTTGGCCGTTGAGCAGAGAGAGAGTCAAGTCAAGACACAGAGGCAATTTTCCGAC	1134
Db	675	TT---CATGAGCCACGCGCAAGAGTGAACGAGATGGACCCGCTGCTGAAAGCCGAAACAG	569
QY	1135	GACCGCTAACCCGCTCCGACAGTCCGCTCAGTTCTCGGACCGGCTCGTGAAGACATGATG	1194
Db	568	GACAGGTACGGCTCTCCGACGTGCGCGCATGGGTGTGGCCGCGAGATCGAGTCAATCCG	509
QY	1195	CACGCTTACTGACTCTCTGCTGTGAAGACACACGACGACCGAACCCGCTTCTGAC	1254
Db	508	GCCGCAACCAAGTCAATCGAGCGGAGAGTCAACTCCGTGAACGACAAACCGGTGATGAC	449
QY	1255	GTCCAGAAACAAGCAGACCGCGACCGGAGGACATTTCCAGGCGTGGGCTGTGATTTGG	1314
Db	448	GTCCACCGCGGCAAGGCGCTCCACAGCGGCACTTCCAGGACACCCCATCGGTGTCTC	389
QY	1315	ATGAGAAAGAACCAAGGCTCGACTGCGCCCTCATCGGCAAGCTCAATTACAGAGTGCAC	1374
Db	388	ATGAGACAACGCGCGTCTCGCCATCGCCAAATCGGCAAGCTCAATGTTCCGCGAGTTCTCC	329
QY	1375	GAGTTGCTCAACGCTGCGATGAACCGGGGCTGCTCTGTGGCTCGTG---CCGAGGAC	1431
Db	328	GAGTTCGTGAACGAGTTCTTACACACAGGGCTGACTTCAACTGGCGGACCGCAAC	269
QY	1432	CCGTGCTCAACTATCACGCGCAAGGCGTTGACATTCACTGCTGTTACGCTTCGAG	1491
Db	268	CCGAGCTTGGACTACGGGTTCAAGGGACCGAAGATGCGCATGCGCTTCTACTGCTCTGAG	209
QY	1492	CTCGGCAACTTGTGCAACCCGGGTCACTACTTGTCTCAGCGCCGACAGATGGGCAACCG	1551
Db	208	CTTCAGTACTCTGCGCAACCCCATCACCAACATGTCTCAGAGCGGAGCGACCAACACAG	149
QY	1552	GCCGCTCAACTGCTGCTCTCATCTCCGCGCGCGCGCATGCGCCAGGSCCAACGACTCTT	1611
Db	148	GAGGTGAACCTCGCTGGGTCTCGTCTGGCCAGGAAGAACCCTGGAAGGCGGTGAGCATCTTC	89
QY	1612	TCTCTCTTCTCGGCTTGGCACTGTACTGACGCTTCAGGCGGTGCACTTCGCGGGAATG	1671
Db	88	AAGCTATGACCTTCACTTACATGTCGCGCTTGGACAGCGCTGATACCTTCCGCACTTC	29
QY	1672	GAG 1674	
Db	28	GAG 26	
RESULT 9			
CB643557/c			
LOCUS			
OSJNB04F02.r OSJNB Oryza sativa (japonica cultivar-group) CDNA			
DEFINITION			
clone OSJNB04F02 3', mRNA sequence.			
ACCESSION			
CB643557			
VERSION			
CB643557.1 GI:29638548			
KEYWORDS			
Oryza sativa (japonica cultivar-group)			

ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaraloideae; Oryzaceae; Oryza.

REFERENCE
1 (bases 1 to 867)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

TITLE
Unpublished (2003)

JOURNAL
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

COMMENT
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: F column: 02
Seq primer: gga aac agc tat gac cat g.

FEATURES
source
1..867
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB04F02"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (che
86061)"

ORIGIN
Query Match 11.0%; Score 265.2; DB 14; Length 867;
Best Local Similarity 59.6%; Pred. No. 1.4e-26;
Matches 504; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

OY 832 CTCGGCCCAAGAGGCTCGTCTGCTCAACGGAAGCGCGCTCCGCTCATGAGCG 891
DB 865 CTGAACCCCAAGAGAGTCTGCCATCGTCATGCGACGTCGTGGGTCCGCGCTGCG 806
OY 892 ACCCTCAGTCTGCACGACTCGACATGCTCTCTCTCGCAGGCTTGACCGCTCTC 951
DB 805 GCCACCGTATGTTGACGCGCAACATCTCTCGCGCTGTCTCGAGGTGCTCTCGCGG 746
OY 952 ACCGTGAGGCGCATGCTCGGCGACGAGGCTGTTCCGCGCGCTTATCAGACGTCTGC 1011
DB 745 TTCTGCGAGGTATGTAACGCGCAAGCGGAGTACACCGACCACTGACCAACACTGAG 686
OY 1012 CGCGCGCACCGCGGACAGGTTCAGAGTCGCGGCGCAATCCGACGCTCTTCGCGCTCG 1071
DB 685 CAC--CACCTCGGTGCTCATGAGCGCGCGCATGAGACATCTCTCGCGGAGC 629
OY 1072 TCGTTTCCGTTGAGCAGCAGAGAGAGTCAAGTCAAGAGCAGCAGGAGCATTTCTTGC 1131
DB 628 TCGTT---CATGAGCCACGCCAAGAGTGAAGAGATGAGACCCGCTCTGAACCCGAG 572
OY 1132 CAGGACCGCTACCGCGTCCGACGCTCGCTCATGTTCTCGCGCGCTCGTGAAGAGCATG 1191
DB 571 CAGGACAGGTACCGGCTCCGACGTCGCGCGAGTGGCTGGCGCGCATGAGGTATC 512
OY 1192 ATGACGCGCTACTCGACTCTCTGCTCGAGAACACAGACGACCGACCGCTCTCTC 1251
DB 511 CGGCGCGCACCAAGTCATCGAGCGCGGAGTCAATCTCGTGAAGACACCGGATGTC 452
OY 1252 GAGCTGAGAGAACAGAGACCGCGCACGCGCGCAATTCGAGCGCTGCTGATTT 1311

DB 451 GAGCTGCACCGCGGCAAGAGCGCTCCACGCGGCAACTTCAGAGGACACCCCATCGGTG 392
OY 1312 TCGATGAGAAAGACAGAGCTCGACTGCGCCCTCATCGGCAAGCTCACTTACGAGATGC 1371
DB 391 TCGATGAGAAAGCGCGCTCGCTCGCATCGCCCAACATCGGCAAGCTCATGTTGCGGCA 332
OY 1372 ACCGATGCTCAACAGCTGCCATTAACCGCGGCTGCTTGTGCTGCTG---CCGAG 1428
DB 331 TCCGAGCTCGTGAACGAGTTTCAACACAGCGGCTGACCTTCAACCTGCGCGAGCGCG 272
OY 1429 GACCGCTGCTCACTATACAGGAGCGGCTTGAGCACTTACATCGCTGTTAGCTTGC 1488
DB 271 AACCGAGCTTGAATAGATAGGAGTTCAAGGACCGAGATCGCATGCTCTTACTGCTCT 212
OY 1489 GAGCTCGGCAACTTGTCCCAACCGCGGTCACTAATTCTGTCAGACCGCGAGAGATGGCAAC 1548
DB 211 GAGCTCGAGTACCTGCGCAACCCCATACCAACCATATCCAGACGCGGAGCAGACAAAC 152
OY 1549 CAGGCGCTCACTGCTGCTCTCATCTCCGCGGCGCGCACTCGCGGAGGCGCAACGAGTC 1608
DB 151 CAGGACGTAACCTGCTGCTGCTCTGCTCGCGCAGAAAGACCTGAGGCGGTGAGATC 92
OY 1609 CTTTCTCTCTTCGCGCTCGACCTGACTGACAGCTTCGAGCGCGCTGCACTTCGCGCG 1668
DB 91 CTCAGCTCATGAGCTCCACTCATCATGTCGCGCTGTGCGAGCGCGTCCGACCGTCCAC 32
OY 1669 ATGAG 1674
DB 31 CTCGAG 26

RESULT 10
CB651503/c 867 bp mRNA linear EST 08-APR-2003
LOCUS OSJNB16H19.1 OSJNB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION OSJNB16H19 3', mRNA sequence.
ACCESSION CB651503
VERSION CB651503.1 GI:29646496
KEYWORDS EST
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaraloideae; Oryzaceae; Oryza.
1 (bases 1 to 867)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

TITLE
Unpublished (2003)

JOURNAL
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

COMMENT
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 16 row: H column: 19
Seq primer: gga aac agc tat gac cat g.

FEATURES
source
1..867
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB16H19"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"

/clone_11b="OSJNEB"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che
 86061)"

ORIGIN

Query Match 11.0%; Score 265.2; DB 14; Length 867;
 Best Local Similarity 59.6%; Pred. No. 1.4e-26;
 Matches 504; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

832 CTCGCGCCGAGAGAGGCTCTGGCTGTGCAACGACGACCGCGCTCCGCTCGATGAGC 891
 865 CTGAACCCCAAGAGAGGCTCTGCGCATCTGCAATGACACGCTCGTGGGTGCGCTCGCG 806
 892 ACCCTGAGTGCACGACTGCAATGCTCTGCTCTCTGCGAGGCTTGAACGCGCTCTC 951
 805 GCCACCGTGAATGTTGACGCGCAACATCTCGCGCTCTGCGAGGTGCTCTGCGCGGTG 746
 952 ACGGTGAGGCGCAGTGGCGCGCAGAGGCTGCTGCGCGCTTATCCAGACGCTTGC 1011
 745 TTCTGCGAGGTGAATGAACGCAAGCGAGTACACCGACACCTTACCAACAGCTGAG 686
 1012 CGCCCGCAGCCCGCGCAGGTGCGAGTGGCGGCAATCCGACGCTCTTTCGCGCTCG 1071
 685 CAC---CACCTGGGTGATCGAGCGCGCGCATATGAGACATCTCGCGCGGAGC 629
 1072 TCGTTTGGCTTGAAGACGAGAGAGGTGCAAGTCAAGTCAAGACGAGAGGCAATTTGCG 1131
 628 TCGTT---CATGAGCCACGCGCAAGAGGTGAACGAGTGAACCCGCTCTGAAGCGAG 572
 1132 CAGAGCCGCTACCCGCTCGCGACGCTGCGCTCAAGTCTTCTGCGCGCTGAGAGACATG 1191
 571 CAGAGCAGGTAAACCGCTCGCGACGCTCGCGCAGTGGCTGCGCGCAGATGAGGTATC 512
 1192 ATGACGCGCTTACTGCACTCTCTGCTGCGAACAACAACGACGACCGACACCCGCTCTC 1251
 511 CGCGCGCCACCAAGTCAATCGAGCGCGAGTCACTCGTGAAGACACACCGGTGATC 452
 1252 GACGTGAGAAACAAGACGAGACCGCGCGCACTTCGAGGCGTGGCTGTCTCAT 1311
 451 GAGTTCACCGCGCAAGGCGCTCCAGCGCGGACATTCAGAGGACACCCCATGAGTGTG 392
 1312 TCGATGAGAAACAAGCTGCGACCTGCGCTCATCGGACGCTCACTTACGACGATGC 1371
 391 TCCATGAGCAACCGCGCTCGCGCATCGCGCAATCGGGAAGCTATTTGCGGAGTTC 332
 1372 ACCGAGTGTCTCAAGCTGCGCATGAACCGCGCGCTGCTGTGCTGCTG---CGAG 1428
 331 TCGAGCTGCTGAACAGTTCATCAACAACGCGGCTGACCTCCACCTGCGCGGACGCGC 272
 1429 GACCGCTGCTCAACTATCATCGGCAAGGCTTGAACATTCATGCTGCTTACGCTTCG 1488
 271 AACCCGAGTGAACGAGGTTCAAGGSCACCGAGATGCGCATGCGCTCTTATGCTCT 212
 1489 GAGCTGAGGCACTTCCCAACCGGCTCACTACTTCTGTCAGCCCGCAGAGATGAGCAAC 1548
 211 GACCTCAAGTACTTCCCAACCGCTCACTCAACCACTGTCAGAGGCGGACACACAAAC 152
 1549 CAGGCGCTCAACTGCTGCTGCTCATCTCGCGCGCGGCACTGCGGAGGCGCAAGAGCTC 1608
 151 CAGGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 92
 1609 CTTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
 91 CTCAGACTCATGACCTTCACCTTACATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32
 1669 ATGAG 1674
 31 CTCGAG 26

RESULT 11
 CB658123/c

LOCUS CB658123 860 bp mRNA linear EST 09-Apr-2003
 DEFINITION OSJNEC1022.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
 ACCESSION CB658123
 VERSION CB658123.1 GI:29661848
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 860)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 CONTACT: Rod Wing
 JOURNAL Arizona Genomics Institute
 COMMENT University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR PRIMERS
 FORWARD: gta aac cga cga cga gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 13 row: 0 column: 22
 Seq primer: gga aac agc tat gac cat g.
 FEATURES
 source
 1..860
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC1022"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_11b="OSJNEC"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 11.0%; Score 265; DB 14; Length 860;
 Best Local Similarity 59.7%; Pred. No. 1.5e-26;
 Matches 502; Conservative 0; Mismatches 330; Indels 9; Gaps 3;

837 CCCGAGAGAGGCTCTGCTGTGTCACGAGACGCGCTCCGCTCGATGCGACCTT 896
 860 CCCAAGGAAGTCTCGCATCTGTCATATGGAAGCTCGTGGGTGCGCGCTCGCGCAC 801
 897 CAGTTCGACGACTGCGACATGCTCTGCTCTCTGCGAGGCTTGAACGCTTCACGAT 956
 800 CGTGAATGTTGAGCGCAACATCTCGCGCTCTGCTGCGAGTGTCTCGCGGCTGTTCTG 741
 957 GAGGCGCATGCTGCGCGCAGAGGCTGCTTTCGCGCGCTTATCCAGACGCTGCGCGCC 1016
 740 CAGAGTATGAACCGCAAGCGGAGTACACCGACACCTTACCAACAGTGAAGAC-- 683
 1017 GACCCCGCGCAGGTGAGGTGCGCGCAACATCCGACGCTCTTTCGCGCTGCTGCTT 1076
 682 -CACCTTGGTGAATGAGGCGCGCGCATATGAGACATCTTTCGCGGAGCTGCTT 624
 1077 TGCGGTGAGCAGAGAGGAGGTCAAGTCAAGACGAGAGGCAATTCCTTTCGCGAGGA 1136
 623 ---CATGAGCCACCGCAAGAGGTGAACGAGTGAACCGCTCTTGAAGCCAGAGAGA 567
 1137 CCGCTACCGGCTCGCAGGTGCGCTCAAGTCTCTGCGCGCGCTGTCGAGAGCATGATGA 1196
 566 CAGGTACCGGCTCGCAGGTGCGCGAGTGTGCTGCGCGCGCATGATGAGGTATTCGCGCG 507

QY 1197 CGCTTACTGACTCTCTGCTCGTAGAACAACAAGACGACGACCAACCCGCTCTCGAGCT 1256
 Db 506 CGCACCAAGATCATGAGAGCGGAGGTCACTCCGTGAACAACAACCGGTGATCGAGCT 447
 QY 1257 CGAAGAACAGAGACCGGCGCAACGCGGCAACTTCAGAGCGTTCGGTGTCTGATTTGAT 1316
 Db 446 CACCCCGGCAAGGCGCTCCAGCGCGCACTTCAGAGGCAACCCCATCGGTGTGTCAT 387
 QY 1317 GAGAAAGACGAGCTTCGACTCGCCTCATCGCAAGCTCAATTCAACGAGTGCACGGA 1376
 Db 386 GAGCAAGCGCCGCTCTGCGATCGCCAAATCGGAGAGCTCATGTTGCGGAGTTCTCCGA 327
 QY 1377 GTTGTCTCAACGCTGCCATGAACCGGCGCTGCTTGTGCTCTGCTG--CCGAGAGACC 1433
 Db 326 GCTCCGGAACGAGTTTACACAACGCGGCTGACCTTCAACCTGGCGGACCCGCAACCC 267
 QY 1434 GTGCTCACTATCAACGAGCGGCTTGAACATTCATTCGCTTACGCTTTCGAGCT 1493
 Db 266 GAGCTTGAATAGAGGTTCAAGGACCGAAGATCGCCATGCGCTCTTCTGCTGAGCT 207
 QY 1494 CGGCACTTGTCCAAACCGGCTCACTAATTGTCAGGCGCGGAGATGGCAACCAAGGC 1553
 Db 206 CCAATCTCTGCGCAACCCCATCAACAACCATGTCAGAGCGGAGAGCAACAACAGGA 147
 QY 1554 CGTCACTGCTGCTGCTCATCTCGCGCGCGCACTGCGGAGCAACGCTCTTTC 1613
 Db 146 GCTGAATCTGCTGAGTCTGCTCTGCGAGAGAACCTTGAAGGCGGTGAGATCTTCA 87
 QY 1614 TCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1673
 Db 86 GCTCATGACTCTGACTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 27
 QY 1674 G 1674
 Db 26 G 26

RESULT 12
 CB651685/ 891 bp mRNA linear EST 08-APR-2003
 DEFINITION OSUNB16L23.r OSUNB Oryza sativa (japonica cultivar-group) cDNA
 OSUNB16L23 3', mRNA sequence.
 CB651685
 CB651685.1 GI:29646678
 EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.
 Jantsuaritayarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 Contact: Rod Wang
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210089, Tucson, AZ
 85721-0089, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gta
 BACKWARD: gga aac agc tat gac cat g
 Plate: 16 row: L column: 23
 Seg primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..891
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"

ORIGIN
 Query Match 10.9%; Score 263.8; DB 14; Length 891;
 Best Local Similarity 59.3%; Pred. No. 2.2e-26;
 Matches 505; Conservative 0; Mismatches 337; Indels 9; Gaps 3;
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSUNB16L23"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSUNB"
 /note="vector: pBluescript II KS+; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (che
 86061)"

QY 827 TCGTCCGCGCCGAGAGAGGCTTCGCTGCTGATCAAGAGAGGCGCTCCGCTCGA 886
 Db 870 TCAGCTGAACCCCAAGAAAGTCTCGCCATCTGTATGAGACGTTCGTGGGCTGGGCG 811
 QY 887 TGGCACTCTAGTGTGACAGACTGCAATGCTCTGCTCTCTGAGGCTTTAGCG 946
 Db 810 TCGCGCCACCGTGAATGTTGACAGCCCAATCTCGCGTCTGTCTGAGGTGCTTCG 751
 QY 947 CTCTACAGGTGAGGCACTGCTGCGCCAGAGGCTGCTTGGCGCCGTTCAATCCAGAG 1006
 Db 750 CGGTGTTCTGCGAGGTATGAACGCAAGCCGAGTACCGACCACTGACCAACAAGC 691
 QY 1007 TCTGCGCCGCGACCCCGGCAAGTGCAGTGCAGCGCAATCCGACGCTCTTCG 1066
 Db 690 TGAAGAC---CACCTGGGTGATGAGGCGCGCATGAGAGCAATCTCTCCG 634
 QY 1067 GCTGCTGTTTCCGTTAGACAGAGAGAGTCAAGTCAAGAGACGACGAGGCAATTC 1126
 Db 633 GAGCTCTGTT---CATGAGCAACGCAAGAAAGTGAAGATGACATCCGCTGTGAAGC 577
 QY 1127 TTGCGAGAGCCGCTACCCGCTCCGACGCTGCTCAATTCCTCGCCGCTCTGAGAG 1186
 Db 576 CGAAGAGAGACGATGACGCTCCGACATGTCGCGCAGTGTGCTGGCCCGACATCGAGG 517
 QY 1187 ACATGATGAGCGCTTACTGACTCTCTGCTGAGAACAAACAAGACCAACCAACCCGC 1246
 Db 516 TCATCGCGCCGCCCAAGTGCATGAGCGCGAGGTAACTCGTGAACAACAACCCG 457
 QY 1247 TCTGCACTGAGAAACAGACGACCGGCAAGCGGCGCACTTCAGAGCGTGGCTGTCT 1306
 Db 456 TGATGACGTCACCGCGCAAGGCGCTCCACGGCGGAACTTCCAGGGCAACCCCATCG 397
 QY 1307 CGATTCGATGAGAAACAGCGCTGCACTGCGCTCTCATGCGCAAGTCAATTACGC 1366
 Db 396 GTGTGTCATGAGAACACCGCGCTGCGCATGCGCAATCGGAGAAAGCTCATGTTGCGGC 337
 QY 1367 AGTGCACGAGTGTGTCACAGCTGCGCATGAACCGGCGCTGCGCTGAGCTGCTG--- 1423
 Db 336 AGTTCCTGAGTCTGTGAACGATTTCTAACAAAGAGGCTAATCTCAACTTGGCGCA 277
 QY 1424 CCGAGACCCGCTGCTCACTATCAAGGCAAGGCTTGAACATTCATCGCTCTTACG 1483
 Db 276 GCGGACCCGAGCTTGAATACGAGTTCAAGGGAACGAGATGCGCATGCTCTACT 217
 QY 1484 CTTGGAAGTGGCACTTGGCAACCCGGTCACTACTTGTCTCAAGCCGCAAGATG 1543
 Db 216 GCTGAGCTCAAGTACCTTGGCAACCCCATACCAATGTCAGAGCGGAGAGC 157
 QY 1544 GCAACGAGCGCTCAACTGCTGCTGCTCATCTCGCGCGCGCACTGCGAGGCAACG 1603
 Db 156 ACAACGAGAGCTGAATCTGCTGAGTCTGCTGCGGCAAGAAACCTTGAAGCGGAG 97
 QY 1604 AGTCTCTTCTCTCTTCTGCTGCTGCACTGCTGATGACGCTCAAGGCGCTGACCTTC 1663
 Db 96 ACATCTCAAGCTCAAGCTCACTCACTCATGCTGCGCTGTGCGAGGCGGACGACCTTC 37
 QY 1664 GGGCATGAG 1674

Db 36 GCCACCTCGAG 26

RESULT 13

CB648470/c 893 bp mRNA linear EST 08-APR-2003
 OSJNB11N20.r OSJNB Oryza sativa (japonica cultivar-group) cDNA
 LOCUS DEFINITION
 CB648470 clone OSJNB11N20 3', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

CB648470.1 GI:29643463
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 893)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 11 row: N column: 20
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers

FEATURES

source
 1..893
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNB11N20"
 /issue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNB"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che
 86061)"

ORIGIN

Query Match 10.9%; Score 263.8; DB 14; Length 893;
 Best Local Similarity 59.3%; Pred.No. 2.2e-26;
 Matches 505; Conservative 0; Mismatches 337; Indels 9; Gaps 3;
 Db 827 TCGTCCGCGCCGAGAGAGGTCTCGTCTGTCACGGAACGCGCTCTCCGCTCGA 886
 870 TACAGCTGAACCCCAAGAGGTCTCGCATGTCATGACACGTCGCGGGGCTCGCGC 811
 QY 887 TGCAGACCTCACTGTCACGACTGACATCTCTCGCTCTCGACGAGCTTTGACGG 946
 Db 810 TCCGCGCACCGTGAATTCGACGCGACATCTCGCTCTGTCGACGCTGCTCGG 751
 QY 947 CTCTACGAGTGAAGGATGTCGTCGACGAGGCTCGTTCGCGGCTTCAACAGACG 1006
 Db 750 CGGTGTTTCGAGGATGATGACGACGACCGAGTACACGACCTGACCCACAGC 691
 QY 1007 TCTGCGCGCCGACCCCGCCAGGTGAGGTGCGCGCAACATCCGACGCTCTTCG 1066
 Db 690 TGAAGAC--CACCTGAGTGCATGAGGCGCGCCATCATGAGACATCTCTCCG 634
 QY 1067 GCTCGCTGTTGCGCTTGAACGAGAGAGTCAAGTCAAGACGACGAGGCAATTC 1126

Db 633 GGAGCTCGT---CATGAGCCACCGCAGAGAGTGAAGAGATGAGACCCGCTGCTAGC 577

QY 1127 TTGCGCAGAGACGCTACCCGCTTCGCGACAGTGGCTTCACTTCTCGCCGCTGAGG 1186

Db 576 CGAAGAGAGACAGTACCGGCTTCGCGACAGTGGCTGAGTGGCTGCGGACATCGAGG 517

QY 1187 ACATGATGACACCTCACTGACTCTCTGCTGCGAGAACACACAGCAGCAGACCCG 1246

Db 516 TCATCCGCGCCCGCACCAAGTTCATTCAGCGCGAGTGAATCTCGGTGAAGACACCCG 457

QY 1247 TCTTGACGTGAGAACAGACGACCGCGCACTTTCAGGCGCTGCTT 1306

Db 456 TATATGACGTTCACCGCGCAAGGCGCTTCCACGCGCACTTCAGGCGACCCCATCG 397

QY 1307 CGATTTCATGAGAGACGAGCGCTGCGACCTCGCTCATCGGACCTCACTTCACG 1366

Db 396 GTGTGTCATGAGACAGCGCCGCTTCGCGCATGCGCAACATGCGCACTATGTCGCGC 337

QY 1367 AGTGCACCGAGTTGCTCAACGCTGCGCATGAACCGCGGCTGCTGCTGCTGCTG-- 1423

Db 336 AGTTCTCGAGCTCGTGAACGAGTTCTAACACAGGAGCTGACCTCCACCTGCGGCA 277

QY 1424 CCGAGAGACCGCTGCTCAACATTCACGCGCAAGGCGTTGACATTCATGCTGCTTACG 1483

Db 276 GCGCAACCGAGCTTGAAGCTGCAAGGCGTTCAGAGGCGACGAGATGCGCATGCTTACT 217

QY 1484 CTTCGAGCTCGCGCCACCTTGCACACCGGCTGACTTCTGTCAGCCGCGAGATGG 1543

Db 216 GCTTGAGCTGAGTACTGCGCAACCGCATTCACCAACATGTCACAGAGCGGAGAGC 157

QY 1544 GCAACGAGCGCTCAACTGCTGCTCATCTCCGCGCGCGACGCTCCGAGCCAGC 1603

Db 156 ACAACAGAGAGCTGAACCTGCTGAGTCTGCTGCTGCGCAGAGAACCTTGAAGCGGTG 97

QY 1604 AGCTCTTCTCTCTCTTCTGCTGCGACCTGATCTCAGCGCTCAGAGCGCTGACCTCC 1663

Db 96 ACATCTCAAGTATATGATCTTCACATGATGCTGCTGCTGCGAGCGGTAGACTTC 37.

QY 1664 GCGCATGAGG 1674

Db 36 GCCACCTCGAG 26

RESULT 14

CB673619/c 868 bp mRNA linear EST 09-APR-2003
 OSJNE08D18.r OSJNE Oryza sativa (japonica cultivar-group) cDNA
 LOCUS DEFINITION
 clone OSJNE08D18 3', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

CB673619.1 GI:29677344
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 868)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aac cga cgg cca gtc

```

BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: D column: 18
Seq primer: gga aac agc tat gac cat g.

```


improving the production of PAL, its analog or another optically active unnatural amino acid having PAL-like structure. The PAL polynucleotides are useful for treating a mammal having a disease, disorder or condition selected from phenylketonuria, cancer, human immunodeficiency virus infection and human cytomegalovirus infection. The present sequence represents a R. graminis PAL polypeptide

Sequence 720 AA;

Query Match 99.9%; Score 3608; DB 5; Length 720;

Best Local Similarity 100.0%; Pred. No. 1,3e-306; Mismatches 0; Indels 0; Gaps 0;

Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIVGELLSD 60
 Db 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIVGELLSD 60
 QY 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDSVDFLKAQLONSVYGYTT 120
 Db 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDSVDFLKAQLONSVYGYTT 120
 QY 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSFSVSGLENTLPLEVVGAMVIRVN 180
 Db 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSFSVSGLENTLPLEVVGAMVIRVN 180
 QY 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSIASGDLSPISYIAGAITGHPDVKN 240
 Db 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSIASGDLSPISYIAGAITGHPDVKN 240
 QY 241 VLEHGTEKIMFARBAISLFGLEAVVLGPKEGGLVNGTAVSMTLSLHSHMLSLLSQ 300
 Db 241 VLEHGTEKIMFARBAISLFGLEAVVLGPKEGGLVNGTAVSMTLSLHSHMLSLLSQ 300
 QY 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFAVEHEEVKXKD 360
 Db 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFAVEHEEVKXKD 360
 QY 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLENNTTTNDPLDVENKQTAHGNFOA 420
 Db 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLENNTTTNDPLDVENKQTAHGNFOA 420
 QY 421 SAVSISMEKTRLALIGKLNFTQCTELNAAAMNGPLSCAAEDPSLNTYHGKLDHIA 480
 Db 421 SAVSISMEKTRLALIGKLNFTQCTELNAAAMNGPLSCAAEDPSLNTYHGKLDHIA 480
 QY 481 AYVSEIGHLANPYTTTVOPEANGQAVNSLALISARPTAEANDVLSLLASHLYCTLOAV 540
 Db 481 AYVSEIGHLANPYTTTVOPEANGQAVNSLALISARPTAEANDVLSLLASHLYCTLOAV 540
 QY 541 DLRAMELDFKQFDPLPTLLOOHGTLGVNATALEVKALNKLRECTTTYYDLEPPWHD 600
 Db 541 DLRAMELDFKQFDPLPTLLOOHGTLGVNATALEVKALNKLRECTTTYYDLEPPWHD 600
 QY 601 AFSYATGTVVEELSSPSANVTLLAVNAMKVASAEKASLTREVRNFMQTPSSQAPAA 660
 Db 601 AFSYATGTVVEELSSPSANVTLLAVNAMKVASAEKASLTREVRNFMQTPSSQAPAA 660
 QY 661 YISPRTRVLXSFVREELGVQARGGVFVGQOETIGSVSRITYAIDGGINHLVKMLA 720
 Db 661 YISPRTRVLXSFVREELGVQARGGVFVGQOETIGSVSRITYAIDGGINHLVKMLA 720

RESULT 2

AAE27938

ID AAE27938 strand; protein; 720 AA.

XX AAE27938;

XX 27-DEC-2002 (first entry)

XX Rhodotorula graminis PAL protein #1.

XX Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;

KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
 KW human cytomegalovirus infection; cytostatic; virucide.

OS Rhodotorula graminis.

FT Key Location/Qualifiers

FT Misc-difference 153

FT /label= Val, Ala

FT /note= "Encoded by GYC"

PN US2002102712-A1.

PD 01-AUG-2002.

PF 24-AUG-2001; 2001US-00939408.

PR 24-JUL-2000; 2000US-00624693.

PR 24-JUL-2001; 2001WO-US023270.

PA (PCBU-) PCBU SERVICES INC.

PI Yoshida RK, Koectera AB;

DR WPI; 2002-690616/74.

DR N-PSDB; AAD45812.

PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase

PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,

PT human immunodeficiency virus or human cytomegalovirus infection.

PS Claim 26; Page 29-31; 74pp; English.

The present invention relates to Yeast (e.g. Rhodotorula) phenylalanine ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding such proteins. PAL sequences are useful for producing L-phenylalanine. They are useful for treating mammals having diseases, disorders or conditions that would benefit from treatment with PAL proteins such as phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomegalovirus (hcmv) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. graminis PAL protein.

Sequence 720 AA;

Query Match 99.9%; Score 3608; DB 5; Length 720;

Best Local Similarity 100.0%; Pred. No. 1,3e-306; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIVGELLSD 60
 Db 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIVGELLSD 60
 QY 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDSVDFLKAQLONSVYGYTT 120
 Db 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDSVDFLKAQLONSVYGYTT 120
 QY 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSFSVSGLENTLPLEVVGAMVIRVN 180
 Db 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSFSVSGLENTLPLEVVGAMVIRVN 180
 QY 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSIASGDLSPISYIAGAITGHPDVKN 240
 Db 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSIASGDLSPISYIAGAITGHPDVKN 240
 QY 241 VLEHGTEKIMFARBAISLFGLEAVVLGPKEGGLVNGTAVSMTLSLHSHMLSLLSQ 300
 Db 241 VLEHGTEKIMFARBAISLFGLEAVVLGPKEGGLVNGTAVSMTLSLHSHMLSLLSQ 300
 QY 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFAVEHEEVKXKD 360
 Db 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFAVEHEEVKXKD 360
 QY 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLENNTTTNDPLDVENKQTAHGNFOA 420


```
DB 361 EGILRQDRYPLRTSPQFLGPLEVDMMAVSTLSIENNTTNDPLDVENKQTHAGNFOA 420
QY 421 SAVSISMKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKGLDIIA 480
DB 421 SAVSISMKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKGLDIIA 480
QY 481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHLYCTLOAV 540
DB 481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHLYCTLOAV 540
QY 541 DIRAMELDFPKQFDPDLLPTLLQOHLGTGLDVNALALEVKALKNKRELEQTTTLEPRMHD 600
DB 541 DIRAMELDFPKQFDPDLLPTLLQOHLGTGLDVNALALEVKALKNKRELEQTTTLEPRMHD 600
QY 601 AFSYATGVVELLSSPSANVTLLTANAMKVASAEKASILTREVNRNFMQTPSSQAPAA 660
DB 601 AFSYATGVVELLSSPSANVTLLTANAMKVASAEKASILTREVNRNFMQTPSSQAPAA 660
QY 661 YLSPTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEAIKGRINHVLVQMLA 720
DB 661 YLSPTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEAIKGRINHVLVQMLA 720
RESULT 3
ABB07694 ID ABB07694 standard; protein; 726 AA.
AC ABB07694;
XX 29-AUG-2003 (revised)
DT 07-AUG-2003 (revised)
DI 10-JUN-2002 (first entry)
XX Rhodotorula pal consensu polypeptide sequence.
XX
XX PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
XX cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
XX consensus; enzyme.
XX Rhodotorula graminis.
XX Rhodotorula mucilaginos.
XX Rhodotorula mucilaginos.
XX Rhodospiridium toruloides.
XX
XX Key Location/Qualifiers
FT Misc-difference 1..726
FT /note= "Xaa is any amino acid; "Xaa" indicates no
FT consensus at that position"
XX
XX WO200208402-A2.
XX
XX 31-JAN-2002.
XX
XX 24-JUL-2001; 2001MO-US023270.
XX
XX 24-JUL-2000; 2000US-00624693.
XX
XX (PCBU-) PCBU SERVICES INC.
XX
XX Yoshida RK, Kootstra AB;
XX
XX WPI; 2002-268973/31.
XX N-PSDB; ABA93244.
XX
XX Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
XX treating mammal having disease or disorder from phenylketonuria, cancer,
XX human immunodeficiency virus infection and human cytomagalovirus
XX infection.
XX
XX Claim 10; Fig 2; 135pp; English.
XX
XX The invention relates to an isolated and purified yeast phenylalanine
```

```
CC ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC host cell comprising the polynucleotide is useful for the production of L
CC -PAL, by adding the polynucleotide or the host cell to a composition
CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC improving the production of PAL, its analog or another optically active
CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC are useful for treating a mammal having a disease, disorder or condition
CC selected from phenylketonuria, cancer, human immunodeficiency virus
CC infection and human cytomagalovirus infection. The present sequence
CC represents a consensus polypeptide sequence of the Rhodotorula PAL.
CC (updated on 07-AUG-2003 to correct OS field.) (updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 726 AA;
```

```
Query Match 76.7%; Score 2770; DB 5; Length 726;
Best Local Similarity 77.7%; Pred. No. 3.7e-23;
Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;
```

```
QY 1 MAPSDLSATTLANGPTNGSHAPTKSAGPTSAIRTPGLDGHAAHQSLFVQLEHLSLD 60
DB 1 MAPSDLSATTSXANGXKNGXHAAXXASXXXXXAXXASXLPPTXQTGLDVEXXLAD 60
QY 61 P-TDQVVELSGYSLTVADVGAARXRRVYQNDDEIRAVDKSVDFLXQLONSVYGV 119
DB 61 PXTDXXELDGYSLTLDGVGAARXRRVYQNDDEIRXIXIDKSVDFLXQLONSVYGV 120
QY 120 TGFSGSADRTEDASISLQKALLEHQLCGVLPFSKDSXGLGLENLPLEVYQAMTIRV 180
DB 121 TGFSGSADRTEDASISLQKALLEHQLCGVLPFSKDSXGLGLENLPLEVYQAMTIRV 180
QY 180 NSLTRGSAVRLVLEALTNFNRITPIVPLRGSISASGDLSPSYIAGAITGHPDVXV 239
DB 181 NSLTRGSAVRLVLEALTNFNRITPIVPLRGSISASGDLSPSYIAGAITGHPDVXV 240
QY 240 HYLHEGTEKIMFARPAISLFGLEAVYLGPKEGLVNGTVASAMATLSLHDSHMLLS 299
DB 241 HYLHEGTEKIMFARPAISLFGLEAVYLGPKEGLVNGTVASAMATLSLHDSHMLLS 300
QY 300 QALTALTYEAMVGGQSFAPFTHDVCZPHQGVVARNIRITLSSGSFAVEHEBEVVKD 359
DB 301 QALTALTYEAMVGGQSFAPFTHDVCZPHQGVVARNIRITLSSGSFAVEHEBEVVKD 360
QY 360 DEGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSIENNTTNDPLDVENKQTHAGN 418
DB 361 DEGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSIENNTTNDPLDVENKQTHAGN 420
QY 419 QASAVSISMKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKGLDII 478
DB 421 QASAVSISMKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKGLDII 480
QY 479 IAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHLYCTLO 538
DB 481 IAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHLYCTLO 540
QY 539 AYDLRAMELDFPKQFDPDLLPTLLQOHLGTGLDVNA---LALVYKALKNKRELEQTT 594
DB 541 AYDLRAMELDFPKQFDPDLLPTLLQOHLGTGLDVNA---LALVYKALKNKRELEQTT 600
QY 595 EPRWHDPSYATGVVELLSSPSANVTLLTANAMKVASAEKASILTREVNRNFMQTPSS 654
DB 601 EPRWHDPSYATGVVELLSSPSANVTLLTANAMKVASAEKASILTREVNRNFMQTPSS 660
QY 655 QAPAHAYLSPTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEAIKGRINH 714
DB 661 SPSALXLYLSPTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEAIKGRINH 720
QY 715 LVQMLA 720
DB 721 LVQMLA 726
```

RESULT 4

```

AAE27942
ID AAE27942 standard; protein; 726 AA.
XX
AC AAE27942;
XX
DT 06-AUG-2003 (revised)
DT 27-DEC-2002 (first entry)
XX
DE Phenylalanine ammonia lyase consensus protein.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytosstatic; virucide.
XX
OS Rhodocorula graminis.
OS Rhodocorula mucilaginosa.
OS Rhodospiridium toruloides.
XX
FH Key Location/Qualifiers
FT Misc-difference /note="Yaa corresponds to unknown amino acid residue"
FT US2002102712-A1.
XX
PD 01-AUG-2002.
XX
PF 24-AUG-2001; 2001US-00939408.
XX
PR 24-JUL-2000; 2000US-00624693.
PR 24-JUL-2001; 2001WO-US023270.
XX
PA (PCBU-) PCBU SERVICES INC.
XX
PI Yoshida RK, Kootstra AB;
XX
DR WPI; 2002-650616/74.
XX
PT Novel isolated and purified Rhodocorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.
XX
PS Claim 26; Page 47-48; 74pp; English.
XX
CC The present invention relates to yeast (e.g. Rhodocorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.
CC They are useful for treating mammals having diseases, disorders or
CC conditions that would benefit from treatment with PAL proteins such as
CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC also used in gene therapy. The present sequence is PAL consensus protein.
CC (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 726 AA;

```

```

Db 181 NSLTRGSAVRLVLEALTNFLNHTIPVLRGTISASGDLSPSYAAAITGHPDSKV 240
Qy 240 HVLHEGTEKIMFARERISLFGAEAVVLGREGGLVNGTAVASAMATLSLDSMTLSL 299
Db 241 HVNHEGTEKIMFARERIALFGLEPVYLGPKESGLVNGTAVASAMATLALDHAMLSL 300
Qy 300 QALTALTVEAMYGQGGSPAFTHDVCRRPHGQVEVARNIRTLGSSFAVEHEEVYKD 359
Db 301 QALTALTVEAMYGAGSFHPFLADVTRPHPTQIEVARNIRTLGSKFAVHHEEVYKD 360
Qy 360 DEGLIKQDPRYPLRTSPQGLPVEDEMMAVSTLSIF-NTTTPDPLDVENKQTAHGNF 418
Db 361 DEGLIKQDPRYPLRTSPQGLPVSMDIHAAVLSLEAGOSTTDNPLDVENKQTHHGNF 420
Qy 419 QASAVSISMEKTRLALAIKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGDH 478
Db 421 QASAVVNTMEKTRLALAIKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGDIA 480
Qy 479 IAAVASELCHLANPVTTVQPAENGNAVNSLALISARTRAEANDVLSLLASHLYCTTQ 538
Db 481 AAAYTSEIGHLANPVTTVQPAENGNAVNSLALISARTRAEANDVLSLLATHTLYCVLQ 540
Qy 539 AVDLRAMELDPFKKQDPPLPTLLOOHLGTDVNA----LALVYKALNRIEQTYYDL 594
Db 541 AVDLRAMELPEFKQDPXPKXXLLKQHPGXHTXXXXKXKXKXKXKXKXKXKXKXKX 600
Qy 595 EPRMHDARFATGIVVELLSPPSANTLTAVNAKVAABKAISLTREVRNRFWQTPSS 654
Db 601 EPRMHDARFATGIVVELLSPPSANTLTAVNAKVAABKAISLTREVRNRFWQTPSS 660
Qy 655 QAPAAVYSPRTRYVSVREELGVQARGGVFVGVQOFTIGSVNSITVATIDGRINRY 714
Db 661 SSPALXVSPRTRYVSVREELGVQARGGVFVGVQOFTIGSVNSITVATIDGRINRY 720
Qy 715 LVKMLA 720
Db 721 LVKMLA 726

RESULT 5
AAE27943
ID AAE27943 standard; protein; 720 AA.
XX
AC AAE27943;
XX
DT 27-DEC-2002 (first entry)
XX
DE Yeast phenylalanine ammonia lyase protein #2.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytosstatic; virucide.
XX
OS Rhodocorula graminis.
XX
FH Key Location/Qualifiers
FT Misc-difference 5 /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT /label=Thr, Ala, Ser
FT FT /note="Encoded by DCV"
FT FT /label=Gly
FT FT /note="Encoded by GGV"
FT FT /label=Ser, Leu

```

FT /note= "Encoded by THG or an in-frame stop codon"
FT Misc-difference 25 /label= Thr, Pro, Ser
FT /note= "Encoded by HCC"
FT Misc-difference 27 /label= Pro, Ser
FT /note= "Encoded by YCG"
FT Misc-difference 28 /label= Ala, Pro
FT /note= "Encoded by SCW"
FT Misc-difference 34 /label= Thr
FT /note= "Encoded by AGH"
FT Misc-difference 36 /label= Arg, Ser
FT /note= "Encoded by MGN"
FT Misc-difference 39 /label= Ala, Pro, Ser
FT /note= "Encoded by BCC"
FT Misc-difference 40 /label= Arg, Gly, Trp
FT /note= "Encoded by BGG"
FT Misc-difference 48 /label= Lys, Thr, Met, Glu, Ala, Val, Gln, Pro, Leu
FT /note= "Encoded by VH3"
FT Misc-difference 54 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 56 /label= Lys, Glu, Gln
FT /note= "Encoded by VAG"
FT Misc-difference 65 /label= Glu, Asp, Val
FT /note= "Encoded by GWS"
FT Misc-difference 66 /label= Ile, Val, Leu
FT /note= "Encoded by VTC"
FT Misc-difference 76 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 87 /label= Thr, Pro, Ser
FT /note= "Encoded by HCB"
FT Misc-difference 93 /label= Asp, Ala
FT /note= "Encoded by GMC"
FT Misc-difference 102 /label= Lys, Asn
FT /note= "Encoded by AAV"
FT Misc-difference 103 /label= Arg, Ser
FT /note= "Encoded by AGB"
FT Misc-difference 109 /label= Thr, Ala, Ser
FT /note= "Encoded by DCB"
FT Misc-difference 112 /label= Asp, His, Tyr
FT /note= "Encoded by BAC"
FT Misc-difference 114 /label= Arg, Ser
FT /note= "Encoded by AGB"
FT Misc-difference 117 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 148 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 150 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 153 /label= Ile, Val, Phe
FT /note= "Encoded by DTC"

FT Misc-difference 154 /label= Glu, Asp
FT /note= "Encoded by GAB"
FT Misc-difference 157 /label= Ser, Gly, Arg
FT /note= "Encoded by VGC"
FT Misc-difference 159 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 183 /label= Thr
FT /note= "Encoded by ACB"
FT Misc-difference 223 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 225 /label= Ser
FT /note= "Encoded by TCB"
FT Misc-difference 237 /label= Ser, Thr, Ile, Gly, Ala, Val, Cys, Phe
FT /note= "Encoded by DVC"
FT Misc-difference 239 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 241 /label= Val
FT /note= "Encoded by GTY"
FT Misc-difference 242 /label= Val, Leu, Phe
FT /note= "Encoded by KTS"
FT Misc-difference 246 /label= Lys, Asn, Thr
FT /note= "Encoded by AMS"
FT Misc-difference 251 /label= Tyr, Ser, Phe
FT /note= "Encoded by THC"
FT Misc-difference 259 /label= Leu, Phe
FT /note= "Encoded by TTD"
FT Misc-difference 305 /label= Leu
FT /note= "Encoded by CTB"
FT Misc-difference 319 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 346 /label= Lys, Arg, Thr, Gln, Pro, Trp, Ser
FT /note= "Encoded by HWG or an in-frame stop codon"
FT Misc-difference 411 /label= Lys, Thr, Met
FT /note= "Encoded by AHG"
FT Misc-difference 421 /label= Thr, Ala, Ser
FT /note= "Encoded by DCC"
FT Misc-difference 457 /label= Leu
FT /note= "Encoded by CTB"
FT Misc-difference 458 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 466 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 487 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 493 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 500 /note= "Encoded by GCH"
FT Misc-difference 518 /label= Thr

```

FT /note= "Encoded by ACB"
FT Misc-difference 536 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 556 /label= Met, Ile, Val, Leu
FT /note= "Encoded by VTB"
FT Misc-difference 557 /label= Ile, Val, Leu
FT /note= "Encoded by VTC"
FT Misc-difference 558 /label= Thr, Ala, Pro
FT /note= "Encoded by VCB"
FT Misc-difference 559 /label= Thr, Ala, Ser
FT /note= "Encoded by DCG"
FT Misc-difference 561 /label= Ile, Leu, Phe
FT /note= "Encoded by HTC"
FT Misc-difference 562 /label= Lys, Glu, Gln
FT /note= "Encoded by VAG"

Query Match 75.6%; Score 2730; DB 5; Length 720;
Best Local Similarity 76.5%; Pred. No. 1.2e-229;
Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

QY 1 MAPSLDGLATTLANGFTNGSHAAPTSAAGPTSALETRPTGLDGHAAHOSOLEIYQELSD 60
DB 1 MAPSDSIATSNXGXNXXHAPKXKXTGATSLXRPXALAPATQXTOLDIXELIAD 60
QY 61 PTDDVVELSGSLTVNDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQIQNSVYGVTT 120
DB 61 PTDDXXELDGYTLTLDVVGAAKGRKRVVQTXDIDIRAKIXXVEFLKXQIXNXXVYVTT 120
QY 121 GFGSSADRTEDAVSLQKALIEHOLCGVPTPTXSFSVSGRGENTLPLEVVRGAMVIRVN 180
DB 121 GFGGSADRTEDDAISIQKALIEHOLCGKXLTSTXSFYLKRGLENSLPLEVVRGAMTIRVN 180
QY 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKVH 240
DB 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTISASGDLSPSYIAAITGHPDXKH 240
QY 241 VHEGTEKIMFARBAISLFGLEAVVLGPRKGLGIWNGTAVASAMATLSHDHMTLSLSQ 300
DB 241 XHHEGTEKIMKAREALALGLEPVLGPRKGLGVNGTAVASAMATLALHDHMTLSLSQ 300
QY 301 ALTALTVEAMVQOQGSFAPFIHDVCRPHRGVAVANIRTLSSGSFAVEHEEVEKVKD 360
DB 301 ALTAATVEAMVGHAGSFHFLHDVTRPHPTQIEVARNIRTLSSGSFAVHHEEVEKVKD 360
QY 361 EGIIRDRERPLRTSPQFLPVEDMMAVSTLSLENNTTTNPILVDENKQTAGNFOA 420
DB 361 EGIIRDRERPLRTSPQFLPVEDMMAVSTLSLENNTTTNPILVDENKQTAGNFOA 420
QY 421 SAVSISMEKTRIALALIGKLNFTOCTELINAAARGLPSCAAEDPSINHYGKGLDHTA 480
DB 421 XAVANMEKTRIALALIGKLNFTOCTELINAAARGLPSCAAEDPSINHYGKGLDHTA 480
QY 481 AVASELGLANFVTTFVOPAEWNOAVNSLALISARPTAEANDVLSLLASHYCTLOAV 540
DB 481 AVTSELXLANFVTTFVOPAEWNOAVNSLALISARPTAEANDVLSLLASHYCTLOAV 540
QY 541 DLRAMELDPKQDFDLPTLLQOHCTGLDYNALALEYKXANRLIOTTTVLEPRMHD 600
DB 541 DLRAMELPKQDFDXXXXLXXQHFGXALDQXELKDKVKKXKXLEQTNISYDLPRMHD 600
QY 601 AFSVATGVVELLSSSPSANVTTLAVANAKVASAEKASLITREVRNRFWOTPSSOAPAH 660
DB 601 AFSFATGVVELLSSSSXXAKVSLAAVNAMKVASAEKASLITRVXRVDFXAPSSSPALX 660
QY 661 YLSPRTVLVSFVREELGVKARGDVPFVGQOETIGSNVSRIVYALIDGGINVLYKMLA 720

```

```

DB 661 YLSPRTVLVSFVREELGVKARGDVPFVGQOETIGSNVSRIVYALIDGGINVLYKMLA 720

RESULT 6
AAP83141
ID AAP83141 standard; protein; 716 AA.
XX
AC AAP83141;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 19-NOV-1990 (first entry)
XX
DE Sequence of L-phenylalanine ammonia lyase (PAL) from Rhodospiridium
DE turoloides.
XX
KM Cinnamic acid; expression vector.
OS Rhodospiridium turoloides.
PN EP278706-A.
PD 17-AUG-1988.
XX
PF 08-FEB-1988; 88EP-00301011.
XX
PR 06-FEB-1987; 87JP-00024705.
PR 18-JUN-1987; 87JP-00152357.
XX
PA (MITR ) MITSUT TOATSU CHEM INC.
XX
PI Fukuhara N, Yoshino S, Sone S, Nakajima Y, Makiguchi N;
XX WPI; 1988-229543/33.
XX
PT Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT having combined promoter of tac promoter and pI promoter for enhanced
XX expression.
XX
PS Example; Page 16-21; 37p; English.
XX
CC Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC for PAL (AAP81101) is described as a reference example. The hybrid
CC plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
CC on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
CC OS field.)
XX
SQ Sequence 716 AA;

Query Match 72.5%; Score 2619; DB 1; Length 716;
Best Local Similarity 72.1%; Pred. No. 6.2e-220;
Matches 523; Conservative 86; Mismatches 102; Indels 14; Gaps 5;

QY 1 MAPSLDGLATTLANGFTNGSHAAPTSAAGPTSALETRPTGLDGHAAHOSOLEIYQELSD 60
DB 1 MAPSLDGLATTLANGFTNGSHAAPTSAAGPTSALETRPTGLDGHAAHOSOLEIYQELSD 60
QY 61 PTDDVVELSGSLTVNDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQIQNSVYGVTT 120
DB 61 PTDDVVELSGSLTVNDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQIQNSVYGVTT 120
QY 121 GFGSSADRTEDAVSLQKALIEHOLCGVPTPTXSFSVSGRGENTLPLEVVRGAMVIRVN 180
DB 121 GFGSSADRTEDAVSLQKALIEHOLCGVPTPTXSFSVSGRGENTLPLEVVRGAMVIRVN 180
QY 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKVH 240
DB 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTISASGDLSPSYIAAITGHPDXKH 240
QY 241 VHEGTEKIMFARBAISLFGLEAVVLGPRKGLGIWNGTAVASAMATLSHDHMTLSLSQ 300
DB 241 VHEGTEKIMFARBAISLFGLEAVVLGPRKGLGIWNGTAVASAMATLSHDHMTLSLSQ 300

```

Db 235 VVHEGKEKILYAREAMALFNLEPVVLGPKESGLGVNGTAVASASMATLALDHAMLSLSQ 294
 301 ALTATVEMVGGQGSFAPFTIHVCRPHGQVEYANIRITLLSGSSPAVHEEYVKKD 360
 295 SLTAMTVAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSRPAVHEEYVKKD 354
 361 EGIIRQDSEYPLRTSPQFGLPEVDMHAYSTLSLE-NTTTDNPDLDEVKQTAHGNFQ 419
 365 EGIIRQDSEYPLRTSPQFGLPEVDMHAYSTLSLE-NTTTDNPDLDEVKQTAHGNFQ 414
 420 ASAVSISMEXTRLALALIGKLNFTQCTELLNANNRGLPSCLAEDPSLNYHGKLDIHI 479
 445 AAANVTMEKTRGLAQIGKLNFTQCTELLNANNRGLPSCLAEDPSLNYHGKLDIHI 474
 480 AAYASELGHANPVTTFYQPAEMGNQAVNSALISARRTAESNDVSLASHLYCTLOA 539
 475 AAYTELGHANPVTTHQPAEMANQAVNSALISARRTAESNDVSLASHLYCTLOA 534
 540 VDLRAMELDFKKQFPLPTLLQCHLG--TGLDV-NALALEVKKALNKRLEQTTVDLE 595
 535 IDLRALIEPFFKKQFQPAVALVSLIDHFGSAMGSLNRLBELVEKVKTLAKRLEQTTVDLV 594
 536 PRMHDAFSYATGTVVELLISSPSANVTITAVANAMKVASAEKALITREVNRRFMQTTSSQ 655
 535 PRMHDAFSYATGTVVELLISSPSANVTITAVANAMKVASAEKALITREVNRRFMQTTSSQ 651
 656 APAHAYLSPRTVLYXSFVREELGVQARGDVFGVQOETIGSNVSRLEYEAIKQGRINHL 715
 652 SPALSYLSPTQILYAFVREELGVQARGDVFLGKQEVTTISNVSKLEYEAIKSGRINNVL 711
 QY 716 VKOLA 720
 Db 712 LKOLA 716
 RESULT 7
 AAB0513
 ID AAB0513 standard; protein; 716 AA.
 AC AAB0513;
 DT 25-MAR-2003 (revised)
 DT 12-NOV-1990 (first entry)
 XX
 DE L-phenylalanine ammonia-lyase.
 XX
 KM L-phenylalanine ammonia-lyase; PAL: foreign gene expression;
 KM culture temperature; expression regulation.
 XX
 OS Rhodosporidium toruloides.
 XX
 PN EP279665-A.
 PD 24-AUG-1988.
 PF 18-FEB-1988; 88EP-00301356.
 PR 19-FEB-1987; 87JP-00034397.
 PR 18-JUN-1987; 87JP-00152359.
 XX
 PA (MITK) MITSUI TOATSU CHEM INC.
 PI Fukuhara N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y,
 DR WPI: 1988-236895/34.
 DR N-PSDB; AAN81116.
 XX
 PT Regulation of expression of foreign gene in Escherichia coli - by
 PT maintaining temp. of culture at 40 deg. C or more to suppress expression.
 XX
 PS Claim 3; Page 15-19; 36pp; English.
 CC The PAL gene is expressed in E.coli carrying a recombinant plasmid.

CC Expression is regulated by maintaining the temperature at at least 40
 CC deg. C to suppress expression. Pal is produced in high concentrations.
 CC See also EP-279664: regulation by sugar component as C source at 0.3% or
 CC more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
 CC 2003 to correct PI field.)
 XX
 SQ Sequence 716 AA:
 Query Match 72.4%; Score 2614; DB 1; Length 716;
 Best local similarity 72.0%; Pred. No. 1,7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
 1 MAPSIDSLATTLANGFTNGSHAPPTSAAGPTSAKRTPLDGHAAHQSLFVQGLSD 60
 1 MAPSIDSLISHSFANG-----VASAKQAVNGASTNLAVAASHLPPTQVTVQVDFVEMKLA 54
 61 PTDVVELSGSYSLVRDVGARGRVRVONDETRAVDKSDVFLKQALQNSVYGVTT 120
 55 PTDVVELSGSYSLVRDVGARGRVRVONDETRAVDKSDVFLKQALQNSVYGVTT 114
 121 GFGGSADRTEDAVSLQKALIEHQLCGVPTFSXSSFSVSGLENTLPLEVVRQAMVIRVN 180
 115 GFGGSADRTEDAVSLQKALIEHQLCGVPTFSXSSFSVSGLENTLPLEVVRQAMVIRVN 174
 161 SLTRGHSARLVLYEALTNPLNRTITVPLRGSISASGSLPSYTAGAITGHDPVKH 240
 175 SLTRGHSARLVLYEALTNPLNRTITVPLRGSISASGSLPSYTAGAITGHDPVKH 234
 241 VHEGTEKIMFARASISLFGLEAVYLGPKEGLGVNGTAVASASMATLALDHAMLSLSQ 300
 235 VVHEGKEKILYAREAMALFNLEPVVLGPKESGLGVNGTAVASASMATLALDHAMLSLSQ 294
 301 ALTATVEMVGGQGSFAPFTIHVCRPHGQVEYANIRITLLSGSSPAVHEEYVKKD 360
 295 SLTAMTVAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSRPAVHEEYVKKD 354
 361 EGIIRQDSEYPLRTSPQFGLPEVDMHAYSTLSLE-NTTTDNPDLDEVKQTAHGNFQ 419
 365 EGIIRQDSEYPLRTSPQFGLPEVDMHAYSTLSLE-NTTTDNPDLDEVKQTAHGNFQ 414
 420 ASAVSISMEXTRLALALIGKLNFTQCTELLNANNRGLPSCLAEDPSLNYHGKLDIHI 479
 445 AAANVTMEKTRGLAQIGKLNFTQCTELLNANNRGLPSCLAEDPSLNYHGKLDIHI 474
 480 AAYASELGHANPVTTFYQPAEMGNQAVNSALISARRTAESNDVSLASHLYCTLOA 539
 475 AAYTELGHANPVTTHQPAEMANQAVNSALISARRTAESNDVSLASHLYCTLOA 534
 540 VDLRAMELDFKKQFPLPTLLQCHLG--TGLDV-NALALEVKKALNKRLEQTTVDLE 595
 535 IDLRALIEPFFKKQFQPAVALVSLIDHFGSAMGSLNRLBELVEKVKTLAKRLEQTTVDLV 594
 536 PRMHDAFSYATGTVVELLISSPSANVTITAVANAMKVASAEKALITREVNRRFMQTTSSQ 655
 535 PRMHDAFSYATGTVVELLISSPSANVTITAVANAMKVASAEKALITREVNRRFMQTTSSQ 651
 656 APAHAYLSPRTVLYXSFVREELGVQARGDVFGVQOETIGSNVSRLEYEAIKQGRINHL 715
 652 SPALSYLSPTQILYAFVREELGVQARGDVFLGKQEVTTISNVSKLEYEAIKSGRINNVL 711
 QY 716 VKOLA 720
 Db 712 LKOLA 716
 RESULT 8
 AAB07693
 ID AAB07693 standard; protein; 716 AA.
 AC AAB07693;
 DT 07-AUG-2003 (revised)
 DT 10-JUN-2002 (first entry)

XX DE R. toruloides PAL polypeptide.
 XX KW PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
 XX KM cyclostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
 XX OS Rhodospiridium toruloides.
 XX PN WO200208402-A2.
 XX PD 31-JAN-2002.
 XX PF 24-JUL-2001; 2001WO-US023270.
 XX PR 24-JUL-2000; 2000US-00624693.
 XX PA (PCBU-) PCBU SERVICES INC.
 XX PI Yoshida RK, Kooester AB,
 XX DR WPI; 2002-268973/31.
 XX DR N-PSDB; ABA95243.
 XX PT Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
 PT treating mammal having disease or disorder from phenylketonuria, cancer,
 PT human immunodeficiency virus infection and human cytomegalovirus
 PT infection.
 XX PS Example 3; Fig 1; 135pp; English.
 XX CC The invention relates to an isolated and purified yeast phenylalanine
 CC ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
 CC host cell comprising the polynucleotide is useful for the production of L
 CC -PAL, by adding the polynucleotide or the host cell to a composition
 CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
 CC improving the production of PAL, its analog or another optically active
 CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
 CC are useful for treating a mammal having a disease, disorder or condition
 CC selected from phenylketonuria, cancer, human immunodeficiency virus
 CC infection and human cytomegalovirus infection. The present sequence
 CC represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
 CC correct OS field.)
 CC XX
 SQ Sequence 716 AA;
 Query Match 72.4%; Score 2614; DB 5; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
 QY 1 MAPSDLSLTTLTANGFTNGSHAPTKSAAGPTSAARRTPGIDGHAAGSOLTEIYQELTSD 60
 DB 1 MAPSDLSISHSPANG-----VASAKQAVNGASTNLAAVAGSHLPTTQVTVYDIVERKMLAA 54
 QY 61 PTDDVVELSGYSLTVRDVVGAAKRGRRVONDEIRARVDSVDFKACLOKNSYGVTT 120
 DB 55 PTDSLTLELDGYSINIGDVVSAARKRPVRVKDSDEIRKIDKSVDFLRQSLMSYGVTT 114
 QY 121 GFGGSAADRTTDAVSLQKALLRHOLCGVTPTSXSSFSVSGLENTLPLEVVRGAMTVRN 180
 DB 115 GFGGSAADRTTDAISLQKALLEHOLCGVLPSSPFSFRGLGLENSTLPLEVVRGAMTVRN 174
 QY 181 SLTRGSAVRLVVLALTNFLNHRITPVPPLRGSTASAGDSPISYAGAITGHPDVKYH 240
 DB 175 SLTRGSAVRLVVLALTNFLNHRITPVPPLRGSTASAGDSPISYAGAITGHPDVKYH 234
 QY 241 VLHGETKIMFAREAISLFGLEAVVLPKESGLGVNGTAVASASWATLSLHDSHMLSLSG 300
 DB 235 VVHEGKEKILYAREMALFENLEPVVLPKESGLGVNGTAVASASWATLSLHDSHMLSLSG 294
 QY 301 ALTALTVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSFVAVHEEVEVYKXD 360
 DB 295 SLTAMTVEAMVGHAQSFPHLDVTRPHPTQIEVAGNIRKLLSGSRFAVHHEEVEVYKXD 354

QY 361 EGI LRQDRYPLRTSPQIFGLPELVEDMMHAYSTLSLE-NTTTDNPILDVENKQTAGGNGFQ 419
 DB 355 EGI LRQDRYPLRTSPQIFGLPELVEDMMHAYSTLSLE-NTTTDNPILDVENKQTAGGNGFQ 414
 QY 420 ASASISMEKRTSLALALGKLNFTQCTELLNAAMRGLPSCILAAEDPSLNGHGLDHI 479
 DB 415 AAIVANTYKTRGLGIAQIGKLNFTQCTELLNAAMRGLPSCILAAEDPSLNGHGLDHI 474
 QY 480 AAYASLGLANLPVTTTQOPAEKQAVNSLALISARTTAPADVLSLASHLYCTLOA 539
 DB 475 AAYTSELGLANLPVTTTQOPAEKQAVNSLALISARTTAPADVLSLASHLYCTLOA 534
 QY 540 VDLRAMELDFKQFPLPPTLLOQLG---GLGV-NALAEVKKALKNTLEQTTTDL 595
 DB 535 IDLRALFEFFKQFPGPALVSLDQFSGAMTGSUNRDELVEKVKTKLAKRIEQTNSYDLV 594
 QY 596 PRMHDAFSYATGTVELLSSPSANVTLTAVANKVSAEKALSLTREVNRFPQTPSSQ 655
 DB 595 PRMHDAFSYATGTVELLSSPSANVTLTAVANKVSAEKALSLTREVNRFPQTPSSQ 651
 QY 656 APAHAYLSPTRTVLSFVREELGVQARRGDVFGVQCEITGSNYSRIYEAKGGRINHV 715
 DB 652 SPALSTLSPTRTVLSFVREELGVQARRGDVFGVQCEITGSNYSRIYEAKGGRINHV 711
 QY 716 VQMLA 720
 DB 712 LQMLA 716
 RESULT 9
 AAE27941
 ID AAE27941 standard; protein; 716 AA.
 AC AAE27941;
 DT 06-AUG-2003 (revised)
 DT 27-DEC-2002 (first entry)
 XX Rhodotorula toruloides PAL protein.
 KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
 KW human cytomegalovirus infection; cyclostatic; virucide.
 XX Rhodospiridium toruloides.
 OS Rhodospiridium toruloides.
 OS US2002102712-A1.
 PD 01-AUG-2002.
 XX 24-AUG-2001; 2001US-00939408.
 PF 24-JUL-2000; 2000US-00624693.
 PR 24-JUL-2001; 2001WO-US023270.
 PA (PCBU-) PCBU SERVICES INC.
 XX Yoshida RK, Kooester AB;
 XX WPI; 2002-690616/74.
 XX N-PSDB; AAD45815.
 PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
 PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
 PT human immunodeficiency virus or human cytomegalovirus infection.
 XX Example 3; Page 43-45; 74pp; English.
 CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC They are useful for treating mammals having diseases, disorders or
 CC conditions that would benefit from treatment with PAL proteins such as

phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomegalovirus (HCMV) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. toruloides PAL protein. (Updated on 06-AUG-2003 to correct OS field.)

Sequence 716 AA:

Query Match 72.4%; Score 2614; DB 5; Length 716;
Best Local Similarity 72.0%; Pred. No. 1.7e-219;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSATTTANGFTNGSHAPTKSAAGPTSAIRTPGIDGHAHQSOLEIVQELLSD 60
1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAAVAGSHLPTQTQVDIVYKMLAA 54
61 PTDVVELSGSYLTVRDVYGAARKGRVVRQNDDEIRAVDXVDFLKACQONSYGVT 120
55 PTDSTLELDGYSLNIGDVYSAARKGRPVKXDSDEIRKIDKSVFELRSQLSMSYGVTT 114
121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSGRGENTLPLEVVRGAMVIRVN 180
115 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSGRGENTLPLEVVRGAMVIRVN 174
181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 240
175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGTSISAGDLSPLSYIAAASGHPDSKVH 234
241 VHEGTEKIMFAREALSIFGLEAVVLGPKREGGLVNGTVASASMATLSLHSHMLSLISQ 300
235 VHEGTEKIMFAREALFNLEPVVLGPKREGGLVNGTVASASMATLSLHSHMLSLISQ 294
301 ALTALIVEAMVQOQGSFAPIFDVCRPHPGQVEYVARNIRITLLSGSSPAVEHEEYKXKD 360
295 SLTAMTVEAMVQOQGSFHPFLHDVTRPHPTQIEVAGNIRKLLGSRFVHHEEYKXKD 354
361 EGLRQDRPLRTSPQFLGPLYEDVMAYSTLSLE-NTTTDNPILDVYENKQTAHGNFQ 419
355 EGLRQDRPLRTSPQFLGPLYEDVMAYSTLSLE-NTTTDNPILDVYENKQTAHGNFQ 414
420 ASAVSISMCKTRLALALIKLNFCTOCTELLNANMRGLPSCIAEDPSLNYHGKLDIHI 479
415 ASAVSISMCKTRLALALIKLNFCTOCTELLNANMRGLPSCIAEDPSLNYHGKLDIHI 474
480 AAYASVLEGLHNPVTTFVQPAEMGQAVNSLALISARTTAENDVLSLLASHYCTLQA 539
475 AAYTSELGLHNPVTTFVQPAEMGQAVNSLALISARTTAENDVLSLLASHYCTLQA 534
540 VDIRAMELDPKQCFDPLPTLLQOHG---TGLDV-NMLALEVKKALKRLEQTTVDLE 595
535 IDIRALEFEFKQCFGPAIVSLIDQHFSGAMTGSNLRDELVEKVNKTLAKRLEQTNISDVL 594
596 PRWDAFSYATGTVVELSSPSANVTLTAVANAKVASAEKAISLIREVNRFWQTPSSQ 655
595 PRWDAFSYATGTVVELSSPSANVTLTAVANAKVASAEKAISLIREVNRFWQTPSSQ 651
656 APAAHAYISPTRLVYFVAREBELGVQARRQDVYVQOQETISGNSRITAEIKDGINHVL 715
652 SPALSYISPTQILVYFVAREBELGVKARRQDVYVQOQETISGNSRITAEIKDGINHVL 711
716 VKMLA 720
712 LKMLA 716

```

RESULT 10
AAE20653 standard; protein; 716 AA.

XX AAE20653;
XX AAE20653;
XX 01-JUL-2002 (first entry)
XX Rhodotorula glutinis phenylalanine ammonia lyase (PAL).
DB

XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
XX phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
XX cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
XX LCP; telecommunication; medical device; aerospace application; enzyme;
XX biocatalyst.

OS Rhodotorula glutinis.

PN MO200210407-A1.

XX 07-FEB-2002.

XX 23-JAN-2001; 2001WO-US002099.

XX 27-JUL-2000; 2000US-00627216.

XX (DURO) DU PONT DE NEMOURS & CO E. I.

XX Gatenby AA, Sariastani FS, Tang X, Qi W, Vannelli T;

XX MPI; 2002-280635/32.

XX N-PSDB; AAD33075.

PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT polypeptide, or mutant TAL polypeptide, useful for producing
PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT hydroxylase.

PS Example 6; Page 92-95; 139pp; English.

XX The invention relates to methods for biological production of para-
XX hydroxycinnamic acid (PHCA). The invention relates to developing of a new
XX biocatalyst for conversion of glucose to PHCA by incorporation of the
XX wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
XX Rhodotorula glutinis into Escherichia coli underlying the ability of the
XX wild-type PAL to convert tyrosine to PHCA. The invention is also directed
XX to developing a new biocatalyst for conversion of glucose to PHCA by
XX incorporation of the wildtype PAL which possesses enhanced tyrosine
XX ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
XX plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
XX Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
XX recombinant techniques. The recombinantly produced PHCA may be used as a
XX monomer for production of liquid crystal polymers (LCP). LCP may be used
XX in electronic connectors and telecommunication and aerospace
XX applications. LCP resistance to sterilizing radiation has also enabled
XX these materials to be used in medical devices as well as chemical, and
XX food packaging applications. The present sequence is Rhodotorula glutinis
XX PAL protein.

XX Sequence 716 AA;

Query Match 72.4%; Score 2614; DB 5; Length 716;
Best Local Similarity 72.0%; Pred. No. 1.7e-219;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSATTTANGFTNGSHAPTKSAAGPTSAIRTPGIDGHAHQSOLEIVQELLSD 60
1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAAVAGSHLPTQTQVDIVYKMLAA 54
61 PTDVVELSGSYLTVRDVYGAARKGRVVRQNDDEIRAVDXVDFLKACQONSYGVT 120
55 PTDSTLELDGYSLNIGDVYSAARKGRPVKXDSDEIRKIDKSVFELRSQLSMSYGVTT 114
121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSGRGENTLPLEVVRGAMVIRVN 180
115 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSGRGENTLPLEVVRGAMVIRVN 174
181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 240
175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGTSISAGDLSPLSYIAAASGHPDSKVH 234
241 VHEGTEKIMFAREALSIFGLEAVVLGPKREGGLVNGTVASASMATLSLHSHMLSLISQ 300

```


RESULT 12
ABG71948 standard; protein; 716 AA.
ID ABG71948
AC ABG71948;
XX 27-JAN-2003 (first entry)
XX R. glutinis wild-type Phenylalanine ammonia-lyase, PAL.
XX Enzyme: phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
XX para-hydroxycinnamic acid; tyrosine ammonia lyase; TAL; tyrosine;
XX cinnamic acid; PHCA; liquid crystal polymer.
XX Rhodotorula glutinis.
XX MO200290523-A2.
XX 14-NOV-2002.
XX 03-MAY-2002; 2002WO-US018551.
XX 04-MAY-2001; 2001US-0288701P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX Qi WM, Saritasani FS, Tang X;
XX WPI; 2003-058928/05.
XX N-PSDB; ABS56530.
XX Novel recombinant host for producing para-hydroxycinnamic acid and
XX tyrosine, comprises gene encoding tyrosine ammonia lyase activity and
XX gene encoding phenylalanine hydroxylase activity.
XX Claim 5; Page 48-50; 69pp; English.
XX The invention relates to a recombinant host comprising at least one gene
XX encoding a tyrosine ammonia lyase (TAL) activity, and at least one gene
XX encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
XX (1) a method for producing para-hydroxycinnamic acid involving: (1)
XX providing the recombinant host; and (2) growing the recombinant organism
XX in the presence of a fermentable carbon substrate and thus producing para
XX -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
XX providing a recombinant organism comprising at least one gene encoding a
XX phenylalanine hydroxylase activity; and (2) growing the recombinant
XX organism in the presence of a fermentable carbon substrate to produce
XX tyrosine. The recombinant host cell is used for producing para-
XX hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
XX production of para-hydroxycinnamic acid (PHCA) which is useful as a
XX monomer for the production of liquid crystal polymer. The present
XX sequence is R. glutinis wild-type Phenylalanine ammonia-lyase, PAL, an
XX enzyme which catalyses the conversion of phenylalanine into cinnamic acid
XX Sequence 716 AA;
SQ

QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSIASAGDLSLSYIAGAITGHPDVKN 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSIASAGDLSLSYIAGAITGHPDVKN 224
QY 241 VLBSTKIMFARPAISLPGLEAVVLGPKKGLGIVNGTAVSASMATLSLHDSMLSLSSQ 300
DB 235 VVHEGKEKILYARFAMALFNLPEVILGPKKGLGIVNGTAVSASMATLSLHDSMLSLSSQ 294
QY 301 ALTALTAVAMVGOQGFAPFTHDVCRRPHGQVEYARIRITLSSGSPFAVHEEEVAKVD 360
DB 295 SLTAVTAVAMVGHAGSHFPLHDVTRPHPTQIEVAGIRKLKSGRRVAVHHEEVAKVD 354
QY 361 EGIIRQDRYPPRTSPQFVGLVEDMKAYSTLSLE-NNITTDNPLDVENKQTAHGNFQ 419
DB 355 EGIIRQDRYPPRTSPQFVGLVEDMKAYSTLSLE-NNITTDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISEKTRRLALIGKINFTQCTELNMAANNRGLPGLCAEDPSLAVHGGKLDIHI 479
DB 415 AAAVANTMEKTRGLAQIGKINFTQCTELNMAANNRGLPGLCAEDPSLAVHGGKLDIHI 474
QY 480 AAVASELGHIANPVTTFVQPAEMGNQAVNSIALISARRTAEANDVLSLLASHLYCTTQA 539
DB 475 AAVTSELGHIANPVTTHQPAEMGNQAVNSIALISARRTAESNDVLSLLASHLYCVLQA 534
QY 540 VDLRAVELDPKKQDPPLPTLLQCHG--TGLDV-NALALEYKALKXLEQTTVDLE 595
DB 535 IDLRALIEFEFFKQFGPAIVSLIDHFGSAMTGNLRELEVEKVKTLAKRLEQNTSYDLV 594
QY 596 PRMHDAPFAGTAVTELLSSPSANVTITAVNAKVAASAKAISLITREVNRRFMOPTPSQ 655
DB 595 PRMHDAPFAGTAVTELLSSPSANVTITAVNAKVAASAKAISLITREVNRRFMOPTPSQ 651
QY 656 APAAHLYSPRTYVLSFVREELGVQARRGDVFGVQOETISNSRITYEAIKORINHL 715
DB 652 SPALSYLSPTQILYAFVREELGVQARRGDVFGVQOETISNSRITYEAIKORINHL 711
QY 716 VKMLA 720
DB 712 VKMLA 716
RESULT 13
ADE36054
ID ADE36054 standard; protein; 716 AA.
XX ADE36054;
XX 29-JAN-2004 (first entry)
XX Rhodosporidium glutinis PAL protein SEQ ID NO:2.
XX DE Rhodosporidium glutinis PAL protein SEQ ID NO:2.
XX KW recombination; nucleic acid template; recombinant polypeptide;
XX altered property; combined gene.
XX OS Rhodosporidium.
XX WO2003072743-A2.
XX 04-SEP-2003.
XX 26-FEB-2003; 2003WO-US005708.
XX 26-FEB-2002; 2002US-0360279P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX Milano J, Tang X;
XX WPI; 2003-721766/68.
XX N-PSDB; ADE36053.
PT Recombination of nucleic acid templates, useful e.g. for producing
enzymes with altered properties, is based on cyclic extension of unpaired

PT primers.
XX Disclousre; SEQ ID NO 2; 481bp; English.
PS
CC The present invention describes a method for the recombination of nucleic
CC acid (NA) templates. The method comprises: (a) providing at least two
CC double-stranded templates (T1, T2) having different 5' and 3' regions in
CC of forward and reverse strands; (b) contacting the templates with at least one each
CC of forward and reverse primers that anneal, respectively, only to the 3'-
CC strand of the antisense strand of T1 and to the 3'-region of the sense
CC strand of T2; (c) extending the primers by no more than 1000 nucleotides
CC (nt); (d) separating extended primers from their templates; (e)
CC reannealing the extended primers to either template; and (f) repeating
CC cycles of steps (b)-(e) until at least one full-length extension product,
CC i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC method starting from at least one each of antisense and sense single-
CC stranded templates; (2) generating a recombinant polypeptide (I) with
CC altered properties by expressing recombination products of the new
CC processes and screening the polypeptides formed for properties different
CC from those of polypeptides from either of the templates; and (3) (1)
CC produced by method (2). The methods are useful for the recombination of
CC nucleic acid templates of interest, and generation of a recombinant
CC polypeptide having altered properties. The method can be used to make
CC combined genes that express proteins e.g. enzymes, cytokines, growth
CC factors, viral proteins or microbial antigens with altered properties,
CC e.g. stability, activity or specificity. The present sequence is used in
CC the exemplification of the present invention.
XX
SQ Sequence 716 AA;
Query Match 72.4%; Score 2614; DB 7; Length 716;
Best Local Similarity 72.0%; Pred. No. 1.7e-219;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MABSLDSLTATLNGFTNGSHAPTKSAAPTSAIRRTGDLGHAHOSOLEIYOELISD 60
Db 1 MABSLDSLSHSFANG-----VASAKQAVNGASTMVAAGSHLPTQYTDYIEKMLAA 54
QY 61 PTDVVVELSGYSTLVNDVVGAKRGRRVQVNDDEIRARVQSDVFLAQIQNSVYGYTT 120
Db 55 PTDSTTELDGYSINLNDVVASAKRGRRVQVNDDEIRARVQSDVFLAQIQNSVYGYTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHOLCGVPTTSXSSFSVSGCHENTLPLEVVRGAMVRVN 180
Db 115 GFGGSADTRTEDAISIQKALIEHOLCGVPTTSXSSFSVSGCHENTLPLEVVRGAMVRVN 174
QY 181 SLRGSASAVRLVYLAITFLNHRITPIVPLRGSISASGDLSPISYAGAITGHPDVKVH 240
Db 175 SLRGSASAVRLVYLAITFLNHRITPIVPLRGSISASGDLSPISYAGAITGHPDVKVH 234
QY 241 VLHEGTEKIMFAREASIFGLBAVVLGPKEGLGVNGTAVASAMATSLIHSHSLISLSQ 300
Db 235 VLHEGTEKILYAREAMALFNLEPVVLGPKEGLGVNGTAVASAMATSLIHSHSLISLSQ 294
QY 301 ALTALIVEAMVQGGSPAFHIDVCRPHPGQVEYARNIRTLSSSPFAVEHEEVKVKXD 360
Db 295 SLTAMTVEAMVGHAGSFHFDVTRPHPTQIEVAGNIRKILEGSRFAVHHEEVKVKXD 354
QY 361 EGIIRGDERYPLRTSPQIFGLIVEDMMHAYSTLSLE-NTTDDNLLLVENKQIRAHGNEFQ 419
Db 355 EGIIRGDERYPLRTSPQIFGLIVEDMMHAYSTLSLE-NTTDDNLLLVENKQIRAHGNEFQ 414
QY 420 AANAVISMEKTRIALALIGKINFTQCTELNLAANMRGLPSCIAEDPSLVNHYGKGLDIH 479
Db 415 AANAVISMEKTRIALALIGKINFTQCTELNLAANMRGLPSCIAEDPSLVNHYGKGLDIH 474
QY 480 AAYASFIHLANPVTTPVOPAEAGNQAANSLAISARTTAANVLSLLASHYKTLQIA 539
Db 475 AAYTSFIHLANPVTTPVOPAEAGNQAANSLAISARTTESNVLSLLASHYKTLQIA 534
QY 540 VDLRAMELDFFKQDPDLPTLLQOCHTG---TGLDV-NALALEVKAALKRLEOTTYDLE 595
Db 535 IDLRAIEFFKQDPFPAIVSLIDGIFGSGAMGSLNRLIEVKVKTAKRLEQTNSEYDLV 594

QY 596 PRWHDAPSVATGTVNVELLSSSPSANVTLTAVNAAMKVAEKAISITREVRNRFQTPSQ 655
Db 595 PRWHDAPSVATGTVNVELLSSSPSANVTLTAVNAAMKVAEKAISITREVRNRFQTPSQ 651
QY 656 APAAVLSPTRTLYSFVARELGVARQDVGVQOETIGSNVRIYEAIKQGRINHL 715
Db 652 SPALSLISRTQTLVAFVARELGVARQDVGVQOETIGSNVRIYEAIKQGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716
RESULT 14
AAE20660
ID AAE20660 standard; protein; 716 AA.
XX
XX AAE20660;
AC
XX
XX 01-JUL-2002 (first entry)
DT
XX
XX Rhodotorula glutinis mutant tyrosine ammonia lyase (TAL), RM120-2 #1.
DE
XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
XX phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
XX cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
XX LCP; telecommunication; medical device; aerospace application; enzyme;
XX biocatalyst; mutant; variant; mutein.
XX
XX Rhodotorula glutinis.
OS
XX Synthetic.
FH
FT Key Location/Qualifiers
FT Misc-difference 198
FT /label= Asp, Asn, Glu, Gln
FT Misc-difference 540
FT /label= Thr, Ala, Ser, Pro, Gly
FT
XX WO200210407-A1.
XX
XX 07-FEB-2002.
PD
XX 23-JAN-2001; 2001WO-US002099.
PP
XX 27-JUL-2000; 2000US-00627216.
PR
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
PA
XX Gatenby AA, Sarlasanti FS, Tang X, Qi W, Vannelli T;
PI WPI; 2002-280635/32.
XX
XX Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT polypeptide, or mutant TAL polypeptide, useful for producing
PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT hydroxylase.
XX
XX Claim 2, Page 126-128, 139pp; English.
PS
XX The invention relates to methods for biological production of para-
CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC biocatalyst for conversion of glucose to PHCA by incorporation of the
CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC to developing a new biocatalyst for conversion of glucose to PHCA by
CC incorporation of the wildtype PAL which possesses enhanced tyrosine
CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC recombinant techniques. The recombinantly produced PHCA may be used as a
CC monomer for production of liquid crystal polymers (LCP). LCP may be used

CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC mutant TAL

XX Sequence 716 AA;

Query Match 72.4%; Score 2613; DB 5; Length 716;

Best Local Similarity 72.0%; Pred. No. 2.1e-219; Mismatches 104; Indels 14; Gaps 5;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAFSLDLSATTLLANGFTNGSHAAPTKSAGPTSAIRRTTGLDGHAAHQSQLEIYQELISD 60
 DB 1 MAFSLDLSISHPANG-----VASAKQAVNGASTYLAAGSHLPTTYQTVQYDIYKMLAA 54
 QY 61 PTDVVELSGYSITVRDVYGAARKGRVAVQNDDEIRARVDKSVDFLKAQONSYYGVT 120
 DB 55 PTDSTLELDGYSINTLDVVAARKGRPVAVKDSDEIRKIDSVFELRSQLSMSYGVTT 114
 QY 121 GFGSADTRTEDAVSLQKALIEHQLCGVPTSSXSFVSGRGLENTLPLEVVRGAMVIRVN 180
 DB 115 GFGSADTRTEDAISTQKALLEHQLCGVLPSPFDSFRJGRLENSLPLEVVRGAMVIRVN 174
 QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSTISASGDISPLSYIAGAITGHPDVYH 240
 DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSTISASGDISPLSYIAAISGHPDSKVH 234
 QY 241 VMEHGEKIMFAEALISLGLENVYIGPEHGIQVNGTVASMTLSLHSHMLSLISQ 300
 DB 235 VMEHGEKILYAEALFNLEFVYIGPEHGIQVNGTVASMTLHLHSHMLSLISQ 294
 QY 301 ALPALTEAMVWGQGSFAPFIHDCRPHGQVEVANIRITLLSGSFVAFVEHEEYKXVD 360
 DB 295 SLTAMTEAMVWAGSFPFIHDCRPHGQVEVANIRITLLSGSFVAFVHHEEYKXVD 354
 QY 361 EGIIRDQRPRLRSPFLGPLYEDMMHANSSTLENTTTNPLLDYENKQTAGNRFQ 419
 DB 355 EGIIRDQRPRLRSPFLGPLYEDMMHANSSTLENTTTNPLLDYENKQTAGNRFQ 414
 QY 420 ASAVSISMEKTRIALIGKLNTOCTELTNAAHNGRLPSCLAADPBLNHYKGLDHI 479
 DB 415 AAVALNTMETRIGLQIGKLNTOCTELTNAAHNGRLPSCLAADPBLNHYKGLDIAA 474
 QY 480 AAYASELGHIANVTTFVQPAEMGNQAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
 DB 475 AAYTSELGHIANVTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
 QY 540 VDIRAMELDPKQFDBLLPTLLOQHIG---TGLDV-NALALEVKAALKRLEQTTYNLE 595
 DB 535 IDIRAAEFERKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKVNKTLLAKRLEQINSYDLV 594
 QY 596 PRMHDAFSYATGVVBLSSPSANVTTLTAVNAAKVASAEKASILTRERVRFMQTPSSQ 655
 DB 595 PRMHDAFSFAGIVVEVLSST---SLSLAIVNAAKVAEASISLTRVRETFVMAASTS 651
 QY 656 APHAHVLSPETRVLYSFFVEELGVARRDDVAVGQOETIGSNVRITPAIKOGFINHVL 715
 DB 652 SPALSYLSPETQTLVYAFVEELGVARRDDVPLKQGEVITIGSNVAKIYEAIKSGINVL 711
 QY 716 VKMLA 720
 DB 712 LKMLA 716

RESULT 15

ABG71951 standard; protein; 716 AA.

XX ABG71951;
 XX DT 27-JAN-2003 (first entry)

DE R. glutinis PAL/TAL mutant #3.

XX Enzyme; phenylalanine ammonia-lyase, PAL; phenylalanine hydroxylase; PAH;
 XX para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
 XX tyrosine; cinnamic acid; PHCA; liquid crystal polymer; mutant; mutein.
 XX Rhodotorula glutinis.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 198 /label= Asp, Asn, Glu, Gln

FT Misc-difference 540 /label= Thr, Ala, Ser, Pro, Gly

PN W0200290523-A2.

PD 14-NOV-2002.

PF 03-MAY-2002; 2002WC-US018551.

PR 04-MAY-2001; 2001US-0288701P.

PP (DUPO) DU PONT DE NEMOURS & CO E I.

XX Q1 WW, Sarlasani FS, Tang X;

DR WPI; 2003-058928/05.

PT Novel recombinant host for producing para-hydroxycinnamic acid and
 PT tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
 PT gene encoding phenylalanine hydroxylase activity.

PS Claim 5; Page 57-59; 69pp; English.

XX The invention relates to a recombinant host comprising at least one gene
 CC encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
 CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
 CC providing the recombinant host; and (2) growing the recombinant organism
 CC in the presence of a fermentable carbon substrate and thus producing para
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
 CC providing a recombinant organism comprising at least one gene encoding a
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant
 CC organism in the presence of a fermentable carbon substrate to produce
 CC tyrosine. The recombinant host cell is used for producing para-
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
 CC monomer for the production of liquid crystal polymer. The present
 CC sequence is a mutant of the R. glutinis Phenylalanine ammonia-lyase, PAL,
 CC which has TAL activity (PAL/TAL).

XX Sequence 716 AA;

Query Match 72.4%; Score 2613; DB 6; Length 716;

Best Local Similarity 72.0%; Pred. No. 2.1e-219; Mismatches 104; Indels 14; Gaps 5;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAFSLDLSATTLLANGFTNGSHAAPTKSAGPTSAIRRTTGLDGHAAHQSQLEIYQELISD 60
 DB 1 MAFSLDLSISHPANG-----VASAKQAVNGASTYLAAGSHLPTTYQTVQYDIYKMLAA 54
 QY 61 PTDVVELSGYSITVRDVYGAARKGRVAVQNDDEIRARVDKSVDFLKAQONSYYGVT 120
 DB 55 PTDSTLELDGYSINTLDVVAARKGRPVAVKDSDEIRKIDSVFELRSQLSMSYGVTT 114
 QY 121 GFGSADTRTEDAVSLQKALIEHQLCGVPTSSXSFVSGRGLENTLPLEVVRGAMVIRVN 180
 DB 115 GFGSADTRTEDAISTQKALLEHQLCGVLPSPFDSFRJGRLENSLPLEVVRGAMVIRVN 174
 QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSTISASGDISPLSYIAGAITGHPDVYH 240
 DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSTISASGDISPLSYIAAISGHPDSKVH 234

QY 241 VHEGTEKIMFARREAIISLFGLEAVVLGPKEGLVNGTAVSASMATLSIJDHSMLSLLSQ 300
Db 235 VHEGKEKILYARERAKLLENLEPVVLGPKEGLVNGTAVSASMATLALHDHMLSLLSQ 294
QY 301 ALTAITYEAMVGOQGSFAPFIHDVCEPHFQGVAVANNIRTLSSGSSFAVEHEEYKXDD 360
Db 295 SLTAMTYEAMVGHAGSFHPLHDVTRHPHQIEVAGNIRKLEGSRFVHHEEYKXDD 354
QY 361 EGIIRORRYPLRTISPOFLGVLVEDMHAISTSLF-NNTTDPPLIDVENKQTAHGNFQ 419
Db 355 EGIIRORRYPLRTISPOFLGVLVEDMHAISTSLF-NNTTDPPLIDVENKQTAHGNFQ 414
QY 420 ASAVSISMETRLALIGLNFQCTELNANRGLPSCLAEDPSLNYHKGJDIHI 479
Db 415 AAAYANTWETKRLGJLQIGLNTFTQITEMLNAGMNGLPSCLAEDPSLNYHKGJDIHI 474
QY 480 AAAYSELGHLANPYTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHIYCTLOA 539
Db 475 AAAYSELGHLANPYTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHIYCTLOA 534
QY 540 VDLAAMELDKQKQDPLFTLLQOHLG---TGIDV-NALALEVKKALNKLEQTTNDLE 595
Db 535 IDLPAKEFEFKQGPALIVSLIDQHGSAANTGSNLRDELVEKVNKTAKRLEQINSYDLV 594
QY 596 PRMHDAFSYATGTVBELSSSPSANYTLTAVNAMKYASAEXAISLREVRANRFMQTPSSQ 655
Db 595 PRMHDAFSYATGTVBELSSSPSANYTLTAVNAMKYASAEXAISLREVRANRFMQTPSSQ 651
QY 656 APAAHYLSPTRYLVSFVREELGVQARQGVFVQOQETIGSNVSRITVBAIKQGRINHL 715
Db 652 SPALSYLSPTRYLVSFVREELGVQARQGVFVQOQETIGSNVSRITVBAIKQGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

Search completed: September 9, 2004, 10:04:00
Job time : 92.6307 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:01:03 ; Search time 23.4025 Seconds
(without alignments)
1588.322 Million cell updates/sec

Title: US-09-939-408A-13

Sequence: 1 MAPSLDLSLTLANGFTNGS.....RIVEAKDGRINVLVLA 720

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3608	99.9	720	4 US-09-624-693A-13	Sequence 13, Appl
2	2770	76.7	726	4 US-09-624-693A-21	Sequence 21, Appl
3	2614	72.4	716	4 US-09-624-693A-19	Sequence 19, Appl
4	2614	72.4	716	4 US-09-627-216A-8	Sequence 8, Appl
5	2614	72.4	716	4 US-09-765-873A-8	Sequence 8, Appl
6	2613	72.4	716	4 US-09-765-873A-35	Sequence 35, Appl
7	2612	72.4	716	4 US-09-627-216A-10	Sequence 10, Appl
8	2612	72.4	716	4 US-09-765-873A-38	Sequence 38, Appl
9	2607	72.2	716	4 US-09-765-873A-36	Sequence 36, Appl
10	2603	72.1	716	4 US-09-765-873A-37	Sequence 37, Appl
11	2603	72.1	716	4 US-09-765-873A-34	Sequence 34, Appl
12	2599	72.0	716	4 US-09-765-873A-33	Sequence 33, Appl
13	2579	71.3	686	4 US-09-765-873A-32	Sequence 32, Appl
14	2573	71.3	686	4 US-09-624-693A-17	Sequence 17, Appl
15	2499.5	69.2	713	4 US-09-624-693A-15	Sequence 15, Appl
16	1346.5	37.3	740	4 US-09-615-192A-337	Sequence 328, App
17	716	19.8	424	4 US-09-615-192A-338	Sequence 328, App
18	524	14.5	525	4 US-09-540-236-285	Sequence 285, Ap
19	508.5	14.1	515	4 US-09-328-352-509	Sequence 509, Ap
20	503.5	13.9	508	4 US-09-489-039A-13575	Sequence 13575, A
21	500.5	13.9	515	4 US-09-282-991A-33046	Sequence 33046, A
22	486	13.5	412	4 US-09-615-192A-330	Sequence 330, App
23	397	11.0	520	4 US-09-252-991A-33049	Sequence 33049, A
24	241	6.7	226	4 US-09-615-192A-337	Sequence 327, App
25	211.5	5.9	97	4 US-09-615-192A-329	Sequence 10, Appl
26	129	3.6	2123	3 US-08-968-685A-10	Sequence 2, Appl
27	127	3.5	2588	3 US-08-936-135-2	Sequence 2, Appl

28	124	3.4	857	4 US-09-252-991A-23956	Sequence 23956, A
29	118.5	3.3	1832	3 US-09-335-409-4	Sequence 4, Appl
30	118.5	3.3	1832	4 US-09-568-102-4	Sequence 4, Appl
31	118.5	3.3	1832	4 US-09-567-969-4	Sequence 4, Appl
32	118.5	3.3	1832	4 US-09-568-480-4	Sequence 4, Appl
33	118.5	3.3	1832	4 US-09-568-486-4	Sequence 4, Appl
34	118.5	3.3	1832	4 US-09-568-472-4	Sequence 4, Appl
35	118.5	3.3	1832	4 US-09-567-899-4	Sequence 53, Appl
36	118	3.3	647	4 US-09-613-303-53	Sequence 331, Appl
37	118	3.3	647	4 US-10-267-311-53	Sequence 4, Appl
38	117	3.2	132	4 US-09-615-192A-331	Sequence 4, Appl
39	117	3.2	1839	2 US-09-404-108-4	Sequence 4, Appl
40	117	3.2	1839	4 US-09-404-108-4	Sequence 4, Appl
41	116.5	3.2	1239	1 US-08-026-1385-3	Sequence 498, Ap
42	116.5	3.2	1589	4 US-09-543-681A-4998	Sequence 30639, A
43	115.5	3.2	366	4 US-09-252-991A-30639	Sequence 54, Appl
44	115.5	3.2	1214	2 US-08-231-193A-54	Sequence 54, Appl
45	115.5	3.2	1214	2 US-08-486-273A-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-09-624-693A-13
Sequence 13, Application US/09624693A
Patent No. 6355468
GENERAL INFORMATION:
APPLICANT: Koestera, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
FILE REFERENCE: 294/9/500NSC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodococcus graminis
US-09-624-693A-13

Query Match	99.9%	Score 3608	DB 4	Length 720
Best Local Similarity	99.9%	Pred. No. 0		
Matches 719	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	MAPSLDLSLTLANGFTNGSHAAPTKSAAGPTSAIRTPGLDGHAAHQSOLFIVEELSD	60	
DB	1	MAPSLDLSLTLANGFTNGSHAAPTKSAAGPTSAIRTPGLDGHAAHQSOLFIVEELSD	60	
QY	61	PTDDVVELSGSYLTVDVVGAAKGRVAVVNDDEIRAVDSVDFLKAQOLNSYGYVT	120	
DB	61	PTDDVVELSGSYLTVDVVGAAKGRVAVVNDDEIRAVDSVDFLKAQOLNSYGYVT	120	
QY	121	GFGSADRTEDAVSLQKALIEHOLCGVTPTSXSSFSVGRGENTLPLEVNGAVIRVN	180	
DB	121	GFGSADRTEDAVSLQKALIEHOLCGVTPTSXSSFSVGRGENTLPLEVNGAVIRVN	180	
QY	181	SLTRGSAVRLVVLALTNPLNHRITPIPLPGSTISASDLSPLSTIGATIGCHDVKN	240	
DB	181	SLTRGSAVRLVVLALTNPLNHRITPIPLPGSTISASDLSPLSTIGATIGCHDVKN	240	
QY	241	VLEHGETKIMFAEATSLPGLFAVVLGPKEGLVNGVTSASMATLSLHDSHMLSLSQ	300	
DB	241	VLEHGETKIMFAEATSLPGLFAVVLGPKEGLVNGVTSASMATLSLHDSHMLSLSQ	300	
QY	301	ALTALTYEAMVGGQGSFAPFIHDVCRPHGQVEVARNITLLSGSFAYEHEEVKXKD	360	
DB	301	ALTALTYEAMVGGQGSFAPFIHDVCRPHGQVEVARNITLLSGSFAYEHEEVKXKD	360	
QY	361	EGILRQRYFLRISPFQFLGIVEDMNAVSTLSLNNTTDNDLVDENKQFAHGGNFQA	420	

```

Db      361  EGILRODRYPLRTSPQPLGLVPEDMHAYSTLSLENNTTDNDLLDVENKQAHGQNFQA 420
Qy      421  SAVSISNEKRLALALIGKLNFTQCTELLNAAANRGPLSCIAEDPSLHYHGKGDIDH 480
Db      421  SAVSISNEKRLALALIGKLNFTQCTELLNAAANRGPLSCIAEDPSLHYHGKGDIDH 480
Qy      481  AYASELGHLANPVTTFVQPAEMGQAVNSIALISARTRAEANDVLSLLASHYCTLOAV 540
Db      481  AYASELGHLANPVTTFVQPAEMGQAVNSIALISARTRAEANDVLSLLASHYCTLOAV 540
Qy      541  DLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVPNALALEVKKALNKRLEQTTTDLDEPRMD 600
Db      541  DLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVPNALALEVKKALNKRLEQTTTDLDEPRMD 600
Qy      601  AFSTATGTVEELLSSSPSANVTTLTAVNAKVASAEKALSTREVRNRFWQTPSSQAPAH 660
Db      601  AFSTATGTVEELLSSSPSANVTTLTAVNAKVASAEKALSTREVRNRFWQTPSSQAPAH 660
Qy      661  YLSRTRVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHVLYKLA 720
Db      661  YLSRTRVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHVLYKLA 720

```

RESULT 2

```

US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocotra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
; OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
; OTHER INFORMATION: Position
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
US-09-624-693A-21

```

Query Match 76.7%; Score 2770; DB 4; Length 726;

Best Local Similarity 77.7%; Pred. No. 1.3e-262;

Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

```

Qy      1  MAPSLDSLATTLANGTNGSHAAPTKSAAGPTSAIRRTPGLDGHAHQOLEIVQELSD 60
Db      1  MAPSLDSLATSXANGXMXHAAXAAXSXXXXXXXXXAXGXLPTTXQTQLDIVEKXLD 60
Qy      61  P-TDIVELESGYSLTRDVVGAARKGRVRYQNDDEIRARVDKSVDFLKQLONSVYGV 119
Db      61  PXTDDXKXELDGYSLTIGDVVGAARKGRVRYXDSDEITKXDKSVEFLRQLOKXSVYGV 120
Qy      120  TGGGSGADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGLENTLPLEVVGAMVIRV 179
Db      121  TGGGSGADTRTEDASLQKALIEHQLCGVLPSTXDSFYLGRGLENSPLEVVGAMVIRV 180
Qy      180  NSITRGSHAVRLVIVLEALTNFLNHRITPVLPRGSIASGDLSPSLVYAGITGHPVYK 239
Db      181  NSLITRGSHAVRLVIVLEALTNFLNHRITPVLPRGSIASGDLSPSLVYAGITGHPVYK 240
Qy      240  HVLEHGETEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299

```

```

Db      241  HVHHEKEXEKMXAREIA1LFGLPEVYLGPKEGLGVNGTAVASAMATLALHDAHMLSL 300
Qy      300  QALTALTVEAMGQQQSFPIHDYCRPRPGQVEVARIIRTLSSSSAVHEEVEVRYKD 359
Db      301  QALTALTVEAMGSHSFPPLHDYCRPRPGQVEVARIIRTLSSSSAVHEEVEVRYKD 360
Qy      360  DEGILRODRYPLRTSPQPLGLVPEDMHAYSTLSLE-NNTTDNDLLDVENKQAHGQNF 418
Db      361  DEGILRODRYPLRTSPQPLGLVSDMHAHAYLSLEAGQSTTNDPLDVENKXHHGQNF 420
Qy      419  QASAVSISNEKRLALALIGKLNFTQCTELLNAAANRGPLSCIAEDPSLHYHGKGDIDH 478
Db      421  QASAVSISNEKRLALALIGKLNFTQCTELLNAAANRGPLSCIAEDPSLHYHGKGDIDH 480
Qy      479  IAAVASELGHANPVTTFVQPAEMGQAVNSIALISARTRAEANDVLSLLASHYCTLO 538
Db      481  IAAVASELGHANPVTTFVQPAEMGQAVNSIALISARTRAEANDVLSLLASHYCTLO 540
Qy      539  AYLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNA---LAEVKKALNKRLEQTTTYDL 594
Db      541  AYLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNA---LAEVKKALNKRLEQTTTYDL 600
Qy      595  EPRMHDAFSTATGTVEELLSSSPSANVTTLTAVNAKVASAEKALSTREVRNRFWQTPSS 654
Db      601  EPRMHDAFSTATGTVEELLSSSPSANVTTLTAVNAKVASAEKALSTREVRNRFWQTPSS 660
Qy      661  QAPAHAYISPRTRVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHV 714
Db      661  QAPAHAYISPRTRVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHV 720
Qy      715  LVKMLA 720
Db      721  LVKMLA 726

```

RESULT 3

```

US-09-624-693A-19
; Sequence 19, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocotra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodococcus toruloides
US-09-624-693A-19

```

Query Match 72.4%; Score 2614; DB 4; Length 716;

Best Local Similarity 72.0%; Pred. No. 2.6e-247;

Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

Qy      1  MAPSLDSLATTLANGTNGSHAAPTKSAAGPTSAIRRTPGLDGHAHQOLEIVQELSD 60
Db      1  MAPSLDSLATSXANGXMXHAAXAAXSXXXXXXXXXAXGXLPTTXQTQLDIVEKXLD 60
Qy      61  P-TDIVELESGYSLTRDVVGAARKGRVRYQNDDEIRARVDKSVDFLKQLONSVYGV 120
Db      61  PXTDDXKXELDGYSLTIGDVVGAARKGRVRYXDSDEITKXDKSVEFLRQLOKXSVYGV 114
Qy      121  TGGGSGADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGLENTLPLEVVGAMVIRV 180
Db      121  TGGGSGADTRTEDAVSLQKALIEHQLCGVLPSTXDSFYLGRGLENSPLEVVGAMVIRV 174
Qy      180  NSITRGSHAVRLVIVLEALTNFLNHRITPVLPRGSIASGDLSPSLVYAGITGHPVYK 239
Db      181  NSLITRGSHAVRLVIVLEALTNFLNHRITPVLPRGSIASGDLSPSLVYAGITGHPVYK 240
Qy      240  HVLEHGETEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299

```



```
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MAPSLDLSLATTLANGFTNGSHAAPTSAAGPTSAALRRTPGLDGHAAQOLEIVQELSD 60
Db 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDDVVELSGYSLVYRDVVGAAKRGRRVYQNDDEIRARVDKSVDFLKAQLONSVGYTT 120
Db 55 PTDDVVELSGYSLVYRDVVGAAKRGRRVYQNDDEIRARVDKSVDFLKAQLONSVGYTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVGAMVIRN 180
Db 115 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVGAMVIRN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKH 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKH 234
QY 241 VLHGETKIMFARBAISLFGLEAVYLGPEKGLGVNGTAVASAMATLSIDSHMLSLSQ 300
Db 235 VVHGEKEXILYAREALFNLBEPVVLGPEKGLGVNGTAVASAMATLSIDSHMLSLSQ 294
QY 301 ALTAALTEAMVGOQGSFAPFIHDVCRPHQGVAVANIRITLLSGSSFAVEHEEVKXKD 360
Db 295 SLTAMTEAMVGHAGSFPHPLHDVTRPHPTQIEVAGNIRKLLEGSRFVHHEEVKXKD 354
QY 361 EGLIKODRYPRLTSPQIFGLVEDMMAHAYTSLSE--NNTTNDNPLDVENKQTAHGNFQ 419
Db 355 EGLIKODRYPRLTSPQIFGLVEDMMAHAYTSLSE--NNTTNDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDIHI 479
Db 415 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDIHI 474
QY 480 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAAMELDPKKQOPDLPTLLOOHLG---TGLDV--NALALEVKKALKRLEQTTTYLE 595
Db 535 IDLRAXEFPFKKQOPDLPTLLOOHLG---TGLDV--NALALEVKKALKRLEQTTTYLE 594
QY 596 PRMHDAFSYATGTVEILLSSPSPANVTITAVANAKVASAKAISLTREVRNRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEILLSSPSPANVTITAVANAKVASAKAISLTREVRNRFWQTPSSQ 651
QY 656 APAHAIVLSPRTVLVSFVREELGVQARRGDVGVQOETIGSNVRIYEAIKQGRINHL 715
Db 652 SPALSYLSPRTQILYAFVREELGVQARRGDVGVQOETIGSNVRIYEAIKQGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 6
US-09-765-873a-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: EC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
```

```
TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X= Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873a-35

Query Match 72.4%; Score 2613; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.2e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSLATTLANGFTNGSHAAPTSAAGPTSAALRRTPGLDGHAAQOLEIVQELSD 60
Db 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDDVVELSGYSLVYRDVVGAAKRGRRVYQNDDEIRARVDKSVDFLKAQLONSVGYTT 120
Db 55 PTDDVVELSGYSLVYRDVVGAAKRGRRVYQNDDEIRARVDKSVDFLKAQLONSVGYTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVGAMVIRN 180
Db 115 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVGAMVIRN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKH 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDVKH 234
QY 241 VLHGETKIMFARBAISLFGLEAVYLGPEKGLGVNGTAVASAMATLSIDSHMLSLSQ 300
Db 235 VVHGEKEXILYAREALFNLBEPVVLGPEKGLGVNGTAVASAMATLSIDSHMLSLSQ 294
QY 301 ALTAALTEAMVGOQGSFAPFIHDVCRPHQGVAVANIRITLLSGSSFAVEHEEVKXKD 360
Db 295 SLTAMTEAMVGHAGSFPHPLHDVTRPHPTQIEVAGNIRKLLEGSRFVHHEEVKXKD 354
QY 361 EGLIKODRYPRLTSPQIFGLVEDMMAHAYTSLSE--NNTTNDNPLDVENKQTAHGNFQ 419
Db 355 EGLIKODRYPRLTSPQIFGLVEDMMAHAYTSLSE--NNTTNDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDIHI 479
Db 415 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDIHI 474
QY 480 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAAMELDPKKQOPDLPTLLOOHLG---TGLDV--NALALEVKKALKRLEQTTTYLE 595
Db 535 IDLRAXEFPFKKQOPDLPTLLOOHLG---TGLDV--NALALEVKKALKRLEQTTTYLE 594
QY 596 PRMHDAFSYATGTVEILLSSPSPANVTITAVANAKVASAKAISLTREVRNRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEILLSSPSPANVTITAVANAKVASAKAISLTREVRNRFWQTPSSQ 651
QY 656 APAHAIVLSPRTVLVSFVREELGVQARRGDVGVQOETIGSNVRIYEAIKQGRINHL 715
Db 652 SPALSYLSPRTQILYAFVREELGVQARRGDVGVQOETIGSNVRIYEAIKQGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 7
US-09-627-216a-10
; Sequence 10, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
```



```
/ APPLICANT: Sariastani, Sima F
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ APPLICANT: Vannelli, Todd
/ APPLICANT: Gatenby, Anthony
/ TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
/ FILE REFERENCE: BC1009 US NA
/ CURRENT APPLICATION NUMBER: US/09/627,216A
/ PRIORITY FILING DATE: 2000-07-27
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: mutant from
US-09-627-216A-10

Query Match      72.4%; Score 2612; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 4e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY      1  MASPDSLATTLANGFTNSHAAPTSAAGPTSAARPTGDLGHAHQSGLEIYVELLSD 60
      1  MASPDSLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTYQYQYDIVEKMLAA 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      61  PTDDVELSGYSLTVADYVGAARKGRVRVONDEIRAVDSYDFELKAQONSYYGTT 120
      55  PTDSTLELDGYSLNIGDVVSAARKGRPVAVKDSDEIRKIDKSVFELNSQSLMSYGYTT 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121  GFGGSAADRTTEDAVSIQKALIEHQLCGVTPTXSFSVGRGIENTLPLEVVRGAMVIRVN 180
      115  GFGGSAADRTTEDAISIQKALLEHQLCGVLPSSFDSFRLGRGLENLPLEVVRGAMTIRVN 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      181  SLTRGSAVRLVYLELTFNLRITPIVPLRGSTIASGDSLPSLYIGATIGHPDVYKH 240
      175  SLTRGSAVRLVYLELTFNLRITPIVPLRGSTIASGDSLPSLYIAAISGHPDSKYH 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241  VHEGTEKIMFAREALSLFGLBAVYLGPKEGLGVNGTAVASAVATLSLHDSHMLSLLSQ 300
      235  VVHEGKEKILYAREAMALFNLBPVVLGPKEGLGVNGTAVASAVATLALHDAHMLSLLSQ 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      301  ALTALVTEAMVGOQGSFAPIFDVCRPHGQVEVARNITTLSSGSAFVHEEVEVKVXD 360
      295  SLTAMTVEAMVAGSGFHFELHDVTRPHPTQLEVAGNIRKLEGSRFVHHEEVEVKVXD 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361  EGIIRDORYPLRTSPQFLGPIVEDMMHAYSTLSLE--NNTTNDNPLLDYENKOTAHGNGFQ 419
      355  EGIIRDORYPLRTSPQWLGPIVSDLIHAAVLTIEAGSTTDNPLIDVENKTSHHGNGFQ 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      420  ASAVSISMEKTRLALALIGKLNFTQCTELNLNAAMNRGIPSCLAADPSLNVHGKGLDHI 479
      415  AAAVANTMEKTRLGLAQIGKLNFTQCTELNLNAGMNRGIPSCLAADPSLSYHCKGLDIAA 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      480  AAASSELGLANVTTFVOPAEKNGQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
      475  AATTSLEGLANVTTHVOPAEAMANDAVNSLALISARPTESNDVLSLLATHLYCTLOA 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      540  VDIRAMELDFKQFDPDLPTLLQOHLG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535  IDLRATEFEFKQFGALIVSLIDQHGSAHMTGSNLRDELVEKVNNTLAKRLEQNTSYDLV 594
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      596  PRMHDAFSYATGVVVELLSSPSANVTTLAVNNAKVAABEKAISLTREVRKRFMOTPSQ 655
      595  PRMHDAFSYATGVVVELLSSPSANVTTLAVNNAKVAABEKAISLTREVRKRFMOTPSQ 651
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      656  APAHAYSPSTRILYAFVREBELGVCAARQGVGVGOQETIGSNVRIEAIKDGKINVL 715
      652  SPALSYSPSTRILYAFVREBELGVCAARQGVGVGOQETIGSNVRIEAIKSGKINVL 711
```

```
QY      716  VKMLA 720
      : : : : :
DB      712  LKMLA 716

RESULT 8
US-09-765-873A-10
/ Sequence 10, Application US/09765873A
/ Patent No. 6521748
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/ FILE REFERENCE: BC1009 US CIP
/ CURRENT APPLICATION NUMBER: US/09/765,873A
/ PRIORITY FILING DATE: 2001-01-19
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodocorulia glutinis
US-09-765-873A-10

Query Match      72.4%; Score 2612; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 4e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY      1  MASPDSLATTLANGFTNSHAAPTSAAGPTSAARPTGDLGHAHQSGLEIYVELLSD 60
      1  MASPDSLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTYQYQYDIVEKMLAA 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      61  PTDDVELSGYSLTVADYVGAARKGRVRVONDEIRAVDSYDFELKAQONSYYGTT 120
      55  PTDSTLELDGYSLNIGDVVSAARKGRPVAVKDSDEIRKIDKSVFELNSQSLMSYGYTT 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121  GFGGSAADRTTEDAVSIQKALIEHQLCGVTPTXSFSVGRGIENTLPLEVVRGAMVIRVN 180
      115  GFGGSAADRTTEDAISIQKALLEHQLCGVLPSSFDSFRLGRGLENLPLEVVRGAMTIRVN 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      181  SLTRGSAVRLVYLELTFNLRITPIVPLRGSTIASGDSLPSLYIGATIGHPDVYKH 240
      175  SLTRGSAVRLVYLELTFNLRITPIVPLRGSTIASGDSLPSLYIAAISGHPDSKYH 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241  VHEGTEKIMFAREALSLFGLBAVYLGPKEGLGVNGTAVASAVATLSLHDSHMLSLLSQ 300
      235  VVHEGKEKILYAREAMALFNLBPVVLGPKEGLGVNGTAVASAVATLALHDAHMLSLLSQ 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      301  ALTALVTEAMVGOQGSFAPIFDVCRPHGQVEVARNITTLSSGSAFVHEEVEVKVXD 360
      295  SLTAMTVEAMVAGSGFHFELHDVTRPHPTQLEVAGNIRKLEGSRFVHHEEVEVKVXD 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361  EGIIRDORYPLRTSPQFLGPIVEDMMHAYSTLSLE--NNTTNDNPLLDYENKOTAHGNGFQ 419
      355  EGIIRDORYPLRTSPQWLGPIVSDLIHAAVLTIEAGSTTDNPLIDVENKTSHHGNGFQ 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      420  ASAVSISMEKTRLALALIGKLNFTQCTELNLNAAMNRGIPSCLAADPSLNVHGKGLDHI 479
      415  AAAVANTMEKTRLGLAQIGKLNFTQCTELNLNAGMNRGIPSCLAADPSLSYHCKGLDIAA 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      480  AAASSELGLANVTTFVOPAEKNGQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
      475  AATTSLEGLANVTTHVOPAEAMANDAVNSLALISARPTESNDVLSLLATHLYCTLOA 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      540  VDIRAMELDFKQFDPDLPTLLQOHLG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535  IDLRATEFEFKQFGALIVSLIDQHGSAHMTGSNLRDELVEKVNNTLAKRLEQNTSYDLV 594
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      596  PRMHDAFSYATGVVVELLSSPSANVTTLAVNNAKVAABEKAISLTREVRKRFMOTPSQ 655
```

|||||:|||||:|:|:|||||:|||||:|:|:|
Db 595 PRMHDAFSPAAGTVEVLST---SLSLAAVNAKVAABESALSTRQVRETFWGAISTS 651
Qy 656 APAAVLSPTRTVLXSPFREELGVQARGVDFVQOETIGSNVSRITYEAIKQGRINVL 715
Db 652 SPALSTLSPRTQILYAFVREELGVKARGDVFILGXQEVITIGSNVSKIYEAIKSGRINVL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 9
US-09-765-873A-38
Sequence 38, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BCI009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 38
LENGTH: 716
TYPE: PRT
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (502)..(502)
OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
NAME/KEY: UNSURE
LOCATION: (540)..(540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match 72.2%; Score 2607; DB 4; Length 716;
Best Local Similarity 71.3%; Pred. No. 1,3e-246;
Matches 521; Conservative 85; Mismatches 105; Indels 14; Gaps 5;

Qy 1 MAPSLDSLTATLTFNGFTNGSHAAPTYSAGPTSAARPTGDLGHAHQSLIEYQELISD 60
Db 1 MAPSLDSLSHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVDIYVKMLAA 54
Qy 61 PTDDVVELSGVSLTVRDVVGAAKGRVAVONDEIRARVDKSVDFLKAQONSVYGVTT 120
Db 55 PTGSTLELDGYSNLNGDVVSAARKGRPVAVKQSDERISKIDSVFELSSQLSMAYGVTT 114
Qy 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTSXSPFSVGRGLENTLPLEVVRGAMVIRVN 180
Db 115 GFGGSADRTEDAVSLQKALIEHQLCGVLPSSFSDFRLGRGLENSLPLEVVRGAMVIRVN 174
Qy 181 SLTRGSAVALVLEALTNFINRITPIVPLRGSTISASGDJSLPLYIGAITGHPDVVH 240
Db 175 SLTRGSAVALVLEALTNFINRITPIVPLRGSTISASGDJSLPLYIAAISGHPDSKVH 234
Qy 241 VLHEGTEKIMFAREASISLFGLEAVVLGPXEGGLVNGTVASASMAITLSLHDSHMLSLISQ 300
Db 235 VVHEGKELIYAREAMALFVLBPVVLGPXEGGLVNGTVASASMAITLALHDAHMLSLISQ 294
Qy 301 ALVALYEAAMVGGQGSFAPFIHDVCRPHQGVAVANRTITLSSGSFAVEHEEYKXVD 360
Db 295 SLTAMVEAMVAGSGFHPFLHDVTRPHPTQILEVAGNIRKLLGSRFVHHEEYKXVD 354
Qy 361 EGIIRQDRYPLRTSPQFLGPLVDDMMHAASITLSLNTTTDNPILLDVENKQTAAGCNFQ 419
Db 355 EGIIRQDRYPLRTSPQFLGPLVDDLIHAHAVITLIEAGSGTTDNPILDVENTSHHGNFQ 414

Qy 420 ASAVSIMEKTELALALIGKLNFTQTELLNAANRGLPSCLAAEDPSLNYHGKGLDIH 479
Db 415 AAAVNTMEKTELGLAQGKLNFTQTEMLVAGNRRGLPCLAAEDPSLSYHGKGLDIAA 474
Qy 480 AAVASLGLANPVTTFPQAPAMGQAVNSLALISARTANAVLSILLASHLYCTLOA 539
Db 475 AAVTSELGLANPVTTTHVQAPAMQAVNSLALISARTESNVLSILLATHLYCVLOA 534
Qy 540 VDLRAMELDFKQFDPDLLPTLLOQHLG---TGLDV-NALALEVKKALNKELEQTTIDLE 535
Db 535 IDLRAMEFDFKQFDPGPAVLSLDQHFSGAMTGSVLRBELVEKVKNTLAKLEQNSVDLV 534
Qy 596 PRMHDAFSPATGTVELLSSPSANVTITAVNAKVAABESALSTRQVRETFWGAISTS 651
Db 595 PRMHDAFSPAAGTVEVLST---SLSLAAVNAKVAABESALSTRQVRETFWGAISTS 651
Qy 656 APAAVLSPTRTVLXSPFREELGVQARGVDFVQOETIGSNVSRITYEAIKQGRINVL 715
Db 652 SPALSTLSPRTQILYAFVREELGVKARGDVFILGXQEVITIGSNVSKIYEAIKSGRINVL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 10
US-09-765-873A-36
Sequence 36, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BCI009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 36
LENGTH: 716
TYPE: PRT
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (181)..(181)
OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
NAME/KEY: UNSURE
LOCATION: (235)..(235)
OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
NAME/KEY: UNSURE
LOCATION: (540)..(540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-36

Query Match 72.1%; Score 2603; DB 4; Length 716;
Best Local Similarity 71.7%; Pred. No. 3,1e-246;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

Qy 1 MAPSLDSLTATLTFNGFTNGSHAAPTYSAGPTSAARPTGDLGHAHQSLIEYQELISD 60
Db 1 MAPSLDSLSHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVDIYVKMLAA 54
Qy 61 PTDDVVELSGVSLTVRDVVGAAKGRVAVONDEIRARVDKSVDFLKAQONSVYGVTT 120
Db 55 PTGSTLELDGYSNLNGDVVSAARKGRPVAVKQSDERISKIDSVFELSSQLSMAYGVTT 114
Qy 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTSXSPFSVGRGLENTLPLEVVRGAMVIRVN 180
Db 115 GFGGSADRTEDAVSLQKALIEHQLCGVLPSSFSDFRLGRGLENSLPLEVVRGAMVIRVN 174

[illegible]

```
/ SEQ ID NO 34
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodotorula glutinis
/ FEATURE:
/   NAME/KEY: UNSURE
/   LOCATION: (126)..(126)
/   OTHER INFORMATION: X= Gly, Ala, Ser, Thr
/   NAME/KEY: UNSURE
/   LOCATION: (138)..(138)
/   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
/   NAME/KEY: UNSURE
/   LOCATION: (540)..(540)
/   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-34
```

Query Match 72.0%; Score 2599; DB 4; Length 716;

Best Local Similarity 71.7%; Pred. No. 7, 6e-246; Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

```
QY 1 MAPSLDLSIATTLANGFTNGSHAAPTSAAGPTSAALRTPGLDGHAAQSLEIVQELISD 60
DB 1 MAPSLDLSISHFANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQYDIVEKMLAA 54
QY 61 PTDDVVELSGSYLTVDVVGAAKGRVRYQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTDTSTLELDGYSINLDVVSAAKGRPVRYKDSDEIRSKIDKSVFELRSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTSXSFSGVSGLENTLPLEVVRGAMVRYN 180
DB 115 GFGGSADRTREKALISIQKALIEHXLGCVLPSSFDRLRGLENSLPLEVVRGAMTIRYN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPVPRLSGSISASGDLSPSYIAGATGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPVPRLSGTISASGDLSPSYIAALSGHPDSKVH 234
QY 241 VLHEGTEKIMFARERASIFGLEAVVLCGRKGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLPEVPLGPRKEGLGVNGTAVASAMATLSLHSHMLSLISQ 294
QY 301 ALTALTYEAAVVGQGSFAPFIHDVCRPHRQVEVANRITRLSGSSFAVEHEEVKXKD 360
DB 295 SLTAMTYEAAVVGAGSFHPFLHDVTRPHPTQIEVAGNIRKLEGSFVHHEEVKXKD 354
QY 361 EGIIRDRYPLRTSPQPLPIVDEMMHAYSTLSLE--NNTTNDPLDJEKQTAHAGNFQ 419
DB 355 EGIIRDRYPLRTSPQPLPIVDSDLIHAHVLITRQGSTTNDPLDJEKQTHSHHGFQ 414
QY 420 ASAVSISMERTLALIGKLNFTQCTELNAAANRGLPSCIAADPSILNHYGKGLDHI 479
DB 415 AAVALNTMEKTRIGLAIQIGKLNFTQCTELNAAANRGLPSCIAADPSILNHYGKGLDIA 474
QY 480 AAYASISGLHANPYTTFVQPAEMGNQAVNSLISARPTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSEIHLANVYTHVQPAEMANOAVNSLISARPTESNDVLSLLATHLYCYVLOA 534
QY 540 VDIRAMELDPKQFDLPTLLQOHLG---TGIDY--NALALEVKKALKRLEQTTVDLE 595
DB 535 IDIRAKEFEFKQFGAIVSLIDQHFSGAMTSNLRDELVEKNKTLAKRLEQTNISDLY 594
QY 596 PRWHDASFYATGTVVELSSSPSANYTLTAVANAKVYASAEKASITREVRNRFQTPSSQ 655
DB 595 PRWHDASFYATGTVVELSSSPSANYTLTAVANAKVYASAEKASITREVRNRFQTPSSQ 651
QY 656 APAAATYSPRTVLYFVREBELGVQARBGDVFGVQOETIGSNVRYEALIKDRINHYL 715
DB 652 SPLASTISPRITQTLIFVREBELGVQARBGDVFGVQOETIGSNVRYEALIKDRINHYL 711
QY 716 VKMLA 720
DB 712 VKMLA 716
```

```
RESULT 13
US-09-765-873A-33
/ Sequence 33, Application US/09765873A
/ Patent No. 6521748
/ GENERAL INFORMATION:
/   APPLICANT: tang, Xiao-Song
/   TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/   FILE REFERENCE: BC1009 US CIP
/   CURRENT APPLICATION NUMBER: US/09/765, 873A
/   CURRENT FILING DATE: 2001-01-19
/   PRIOR APPLICATION NUMBER: US 09/627, 216
/   PRIOR FILING DATE: 2000-07-27
/   PRIOR APPLICATION NUMBER: US 60/147, 719
/   PRIOR FILING DATE: 1999-08-06
/   NUMBER OF SEQ ID NOS: 38
/   SOFTWARE: Microsoft Office 97
/ SEQ ID NO 33
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodotorula glutinis
/ FEATURE:
/   NAME/KEY: UNSURE
/   LOCATION: (126)..(126)
/   OTHER INFORMATION: X= Gly, Ser, Ala, Thr
/   NAME/KEY: UNSURE
/   LOCATION: (138)..(138)
/   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
/   NAME/KEY: UNSURE
/   LOCATION: (149)..(149)
/   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
/   NAME/KEY: UNSURE
/   LOCATION: (181)..(181)
/   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
/   NAME/KEY: UNSURE
/   LOCATION: (198)..(198)
/   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
/   NAME/KEY: UNSURE
/   LOCATION: (202)..(202)
/   OTHER INFORMATION: X= Val, Met, Leu, Cys
/   NAME/KEY: UNSURE
/   LOCATION: (235)..(235)
/   OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
/   NAME/KEY: UNSURE
/   LOCATION: (502)..(502)
/   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
/   LOCATION: (540)..(540)
/   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-33
```

Query Match 71.4%; Score 2579; DB 4; Length 716;

Best Local Similarity 71.0%; Pred. No. 7e-244; Matches 515; Conservative 86; Mismatches 110; Indels 14; Gaps 5;

```
QY 1 MAPSLDLSIATTLANGFTNGSHAAPTSAAGPTSAALRTPGLDGHAAQSLEIVQELISD 60
DB 1 MAPSLDLSISHFANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQYDIVEKMLAA 54
QY 61 PTDDVVELSGSYLTVDVVGAAKGRVRYQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTDTSTLELDGYSINLDVVSAAKGRPVRYKDSDEIRSKIDKSVFELRSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTSXSFSGVSGLENTLPLEVVRGAMVRYN 180
DB 115 GFGGSADRTREKALISIQKALIEHXLGCVLPSSFDRLRGLENSLPLEVVRGAMTIRYN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPVPRLSGSISASGDLSPSYIAGATGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPVPRLSGTISASGDLSPSYIAALSGHPDSKVH 234
QY 241 VLHEGTEKIMFARERASIFGLEAVVLCGRKGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLPEVPLGPRKEGLGVNGTAVASAMATLSLHSHMLSLISQ 294
```


QY 173 GAVYIRVNSLTRGSHAVRVVLEALTNFLNHRITPIVPLRGSIASGDLSPUSYTAGAIT 232
 DB 173 GAMTIRVNSLTRGSHAVRVVLEALTNFLNHRITPIVPLRGSIASGDLSPUSYTAGAIT 232
 QY 233 GHPDVKVHVLHEGTEKIMFARERAIISFGLERAVVLGPKGGLGVNGTAVASAMATLSLHDS 292
 DB 233 GHPDSKVHV--DG--KIMSAGEAIKLGLOPVVLGPKGGLGVNGTAVASAMATLSLHDS 288
 QY 293 HMLSLLOALTALVEAMNGOQSFAPFIHDVGRPHPGQVEVARNIRTLSSGSSFAVEHE 352
 DB 289 HVLSLLOALTALVEAMNGHAGSFHPLHDVTRPHPTQIEVARNIRTLSSGSSFAVEHE 348
 QY 353 EEVKVKODEGILRODRYPLRTSPQFLGPLVEDMMHAYSTLSLE--NNTTNDNPLDENKQ 411
 DB 349 TEVVKVKODEGILRODRYPLRTSPQFLGPLVEDMMHAYSTLSLE--NNTTNDNPLDENKQ 408
 QY 412 THAGGNFQASAVSISNEKTRIALALIGKLNFTQCTELNNAAMRGIPSCLAEDPSLNYH 471
 DB 409 THHGAFMASSVGNTEKTRIAVALMGKVSFTQITEMLNAGMNRALPSCLAEDPSLNYH 468
 QY 472 GKGLDIHIAVASELGHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLAS 531
 DB 469 CKGLDIHIAVASELGHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLAT 528
 QY 532 HLYCTLOAVDIRAVELDFFKQFDPLPTLLIQHLGTGLDVNALAL----EYKALNRRL 586
 DB 529 HLYCVLQAVDIRAVELFEHTKAFEPMTVELLKQHFQ-----ALATAEVEDKVRKSIYKRL 582
 QY 587 ECTTYDLEPRMHDAPFANGTVVELSSPSANVTLLTAVNAMKVASAEKASLREVRN 646
 DB 583 QONNSYDLBOQMHDTFSAVAGAVEALAQ--EVSLSLNAKVAQAEKALATLRSVRD 639
 QY 647 RFWQTPSSQAPAHAYLSPTTRVLYSFVREBELGVQARRGDVFWVOQETIGSNVSRIVEAI 706
 DB 640 SFWAAPSSSSPALKYLSPTRVLYSFVREBVGKARBGDVLGKQEVITIGTNVSRIVEAI 699
 QY 707 KQGRINHVLVKMLA 720
 DB 700 KSGCIAPVLVKMA 713

Search completed: September 9, 2004, 10:09:38
 Job time : 26.4025 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:07:54 ; Search time 92.6141 Seconds
(without alignments)

2493.103 Million cell updates/sec

Title: US-09-939-408a-13

Perfect score: 3610
Sequence: 1 MAPSLDLSLATTLANGFTNGS.....RIYKAIKXGRINHYLVKMLA 720

Scoring table: ELOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3608	99.9	720	9 US-09-939-408a-13	Sequence 13, Appl
2	2770	76.7	726	9 US-09-939-408a-21	Sequence 21, Appl
3	2730	75.6	720	9 US-09-939-408a-30	Sequence 30, Appl
4	2614	72.4	716	9 US-09-765-873a-8	Sequence 8, Appl
5	2614	72.4	716	9 US-09-939-408a-19	Sequence 19, Appl
6	2614	72.4	716	14 US-10-138-970A-4	Sequence 4, Appl
7	2614	72.4	716	14 US-10-188-523B-8	Sequence 8, Appl
8	2614	72.4	716	15 US-10-374-366-2	Sequence 2, Appl
9	2614	72.4	716	15 US-10-439-478-2	Sequence 35, Appl
10	2613	72.4	716	9 US-09-765-873a-35	Sequence 19, Appl
11	2613	72.4	716	14 US-10-138-970A-19	Sequence 4, Appl
12	2613	72.4	716	14 US-10-188-523B-15	Sequence 8, Appl
13	2613	72.4	716	15 US-10-439-478-15	Sequence 35, Appl
14	2612	72.4	716	9 US-09-765-873a-10	Sequence 10, Appl
15	2612	72.4	716	14 US-10-138-970A-24	Sequence 24, Appl

16	2612	72.4	716	14 US-10-188-523B-10	Sequence 10, Appl
17	2612	72.4	716	15 US-10-374-366-4	Sequence 4, Appl
18	2609	72.3	716	15 US-10-374-366-24	Sequence 24, Appl
19	2608	72.2	716	15 US-10-374-366-22	Sequence 22, Appl
20	2607	72.2	716	9 US-09-765-873a-38	Sequence 38, Appl
21	2607	72.2	716	14 US-10-138-970A-22	Sequence 22, Appl
22	2607	72.2	716	14 US-10-188-523B-38	Sequence 38, Appl
23	2607	72.2	716	15 US-10-374-366-16	Sequence 16, Appl
24	2607	72.2	716	15 US-10-439-478-18	Sequence 18, Appl
25	2606	72.2	716	15 US-10-374-366-8	Sequence 8, Appl
26	2605	72.2	716	15 US-10-374-366-10	Sequence 10, Appl
27	2604	72.1	716	15 US-10-374-366-14	Sequence 14, Appl
28	2603	72.1	716	9 US-09-765-873a-36	Sequence 36, Appl
29	2603	72.1	716	9 US-09-765-873a-37	Sequence 37, Appl
30	2603	72.1	716	14 US-10-138-970A-20	Sequence 20, Appl
31	2603	72.1	716	14 US-10-138-970A-21	Sequence 21, Appl
32	2603	72.1	716	14 US-10-188-523B-36	Sequence 36, Appl
33	2603	72.1	716	14 US-10-188-523B-37	Sequence 37, Appl
34	2603	72.1	716	15 US-10-439-478-16	Sequence 16, Appl
35	2601	72.1	716	15 US-10-439-478-17	Sequence 17, Appl
36	2601	72.0	716	15 US-10-374-366-12	Sequence 12, Appl
37	2599	72.0	716	9 US-09-765-873a-34	Sequence 34, Appl
38	2599	72.0	716	14 US-10-138-970A-18	Sequence 18, Appl
39	2599	72.0	716	14 US-10-188-523B-34	Sequence 34, Appl
40	2599	72.0	716	15 US-10-374-366-18	Sequence 18, Appl
41	2599	72.0	716	15 US-10-439-478-14	Sequence 14, Appl
42	2598	72.0	716	15 US-10-374-366-6	Sequence 6, Appl
43	2598	72.0	716	15 US-10-374-366-20	Sequence 20, Appl
44	2579	71.4	716	9 US-09-765-873a-33	Sequence 33, Appl
45	2579	71.4	716	14 US-10-138-970A-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-939-408a-13
Sequence 13, Application US/09939408A
Patient No. US20020102712H1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
APPLICANT: Koestera, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939,408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodocorula graminis
FEATURE:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408a-13

Query Match 99.9% ; Score 3608 ; DB 9 ; Length 720 ;
Best Local Similarity 100.0% ; Pred. No. 0 ;
Matches 720 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;

QY 1 MAPSLDLSLATTLANGFTNGSHAFTKSAAGPFSALRRTPGDGHAAHOSQLEIVGELLSD 60
DB 1 MAPSLDLSLATTLANGFTNGSHAFTKSAAGPFSALRRTPGDGHAAHOSQLEIVGELLSD 60
QY 61 PTDVVVLSGYSLTVADVGAARGRVRVQVNDDEIRARVDKSVDFLKAQJLONSIVGVTT 120

```

Db      61 PTDVVELSGYSLVRVVGAAKRGVRVQNDDEIRARVDSVDFKQLONSVYGVTT 120
Qy      121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSQVGLNTLPLEVVGAMVIRN 180
Db      121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSQVGLNTLPLEVVGAMVIRN 180
Qy      181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPLYAGAITGHPDVX 240
Db      181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPLYAGAITGHPDVX 240
Qy      241 VLHGETEKIMFARBAISLFGLEAVVLGPKGGLVNGTAVSASMATLSLHDSHMLSL 300
Db      241 VLHGETEKIMFARBAISLFGLEAVVLGPKGGLVNGTAVSASMATLSLHDSHMLSL 300
Qy      301 ALTALTVEAMVGQGSFAPFIHDVCRPHQVAVANIRTLISGSSFAVEHEEVKVD 360
Db      301 ALTALTVEAMVGQGSFAPFIHDVCRPHQVAVANIRTLISGSSFAVEHEEVKVD 360
Qy      361 EGLRQDRYPLRTSPQPLGVLVEDMMHAYSTLSLENNTTDNDPLDVENKQTAHGNF 420
Db      361 EGLRQDRYPLRTSPQPLGVLVEDMMHAYSTLSLENNTTDNDPLDVENKQTAHGNF 420
Qy      421 SAVSISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGLD 480
Db      421 SAVSISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGLD 480
Qy      481 AYASELGHANPVTTFVQPAEMGNQAVNSIALISARTRAEADVLSLLASHLYCTLO 540
Db      481 AYASELGHANPVTTFVQPAEMGNQAVNSIALISARTRAEADVLSLLASHLYCTLO 540
Qy      541 DLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTVDLEPR 600
Db      541 DLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTVDLEPR 600
Qy      601 AFSAVATGTVVELSSPSANVTLTAVANMKVASAEKASLSTREVNRFWQTPSSQA 660
Db      601 AFSAVATGTVVELSSPSANVTLTAVANMKVASAEKASLSTREVNRFWQTPSSQA 660
Qy      661 YLSRTRVXSVFVREELGVQARGDVFGVQOETIGSNVSRITYEAIKQGRINHY 720
Db      661 YLSRTRVXSVFVREELGVQARGDVFGVQOETIGSNVSRITYEAIKQGRINHY 720

```

RESULT 2

```

; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Xaa"means any amino acid; "Xaa"means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19

```

US-09-939-408a-21

Query Match 76.7%; Score 2770; DB 9; Length 726;
 Best Local Similarity 77.7%; Pred. No. 5,8e-243;
 Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

```

Qy      1 MAPSLDLSLTTLANGFNGNSHAPFKSAAGPFSALRPPGLDGHAAHOSQLEIYQELSLD 60
Db      1 MAPSLDLSLTTLANGFNGNSHAPFKSAAGPFSALRPPGLDGHAAHOSQLEIYQELSLD 60
Qy      61 P-TDVVELSGYSLFVRDVVGAARKGRVRVQNDDEIRARVDSVDFKQLONSVYGVTT 119
Db      61 P-TDVVELSGYSLFVRDVVGAARKGRVRVQNDDEIRARVDSVDFKQLONSVYGVTT 119
Qy      120 TGFSGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSQVGLNTLPLEVVGAMVIRN 179
Db      120 TGFSGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSQVGLNTLPLEVVGAMVIRN 180
Qy      180 NSLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPLYAGAITGHPDVX 239
Db      180 NSLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPLYAGAITGHPDVX 240
Qy      240 VLHGETEKIMFARBAISLFGLEAVVLGPKGGLVNGTAVSASMATLSLHDSHMLSL 299
Db      240 VLHGETEKIMFARBAISLFGLEAVVLGPKGGLVNGTAVSASMATLSLHDSHMLSL 300
Qy      300 QALTALTVEAMVGQGSFAPFIHDVCRPHQVAVANIRTLISGSSFAVEHEEVKVD 359
Db      300 QALTALTVEAMVGQGSFAPFIHDVCRPHQVAVANIRTLISGSSFAVEHEEVKVD 360
Qy      360 DEGLRQDRYPLRTSPQPLGVLVEDMMHAYSTLSLENNTTDNDPLDVENKQTAHGNF 418
Db      360 DEGLRQDRYPLRTSPQPLGVLVEDMMHAYSTLSLENNTTDNDPLDVENKQTAHGNF 420
Qy      419 QASAVSISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGLD 478
Db      419 QASAVSISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGLD 480
Qy      479 IAAVASELGHANPVTTFVQPAEMGNQAVNSIALISARTRAEADVLSLLASHLYCTLO 538
Db      479 IAAVASELGHANPVTTFVQPAEMGNQAVNSIALISARTRAEADVLSLLASHLYCTLO 540
Qy      539 AYDLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTVD 594
Db      539 AYDLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTVD 600
Qy      595 EPRMHDASVATGTVVELSSPSANVTLTAVANMKVASAEKASLSTREVNRFWQTPSS 654
Db      595 EPRMHDASVATGTVVELSSPSANVTLTAVANMKVASAEKASLSTREVNRFWQTPSS 660
Qy      655 QAPAAVYLSPTRVXSVFVREELGVQARGDVFGVQOETIGSNVSRITYEAIKQGRINHY 714
Db      655 QAPAAVYLSPTRVXSVFVREELGVQARGDVFGVQOETIGSNVSRITYEAIKQGRINHY 720
Qy      715 LVKMLA 720
Db      721 LVKMLA 726

```

RESULT 3

```

; Sequence 30, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693

```



```
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ PRIOR FILING DATE: 2001-07-24
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 30
/ LENGTH: 720
/ TYPE: PR
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Preferred
/ OTHER INFORMATION: Theoretical sequence based in part on SEQ ID NO:20
/ NAME/KEY: misc_feature
/ LOCATION: (5)..(5)
/ OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (12)..(12)
/ OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (16)..(16)
/ OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (17)..(17)
/ OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (19)..(19)
/ OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
/ NAME/KEY: misc_feature
/ LOCATION: (20)..(20)
/ OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
/ NAME/KEY: misc_feature
/ LOCATION: (25)..(25)
/ OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (27)..(27)
/ OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (28)..(28)
/ OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
/ NAME/KEY: misc_feature
/ LOCATION: (34)..(34)
/ OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
/ NAME/KEY: misc_feature
/ LOCATION: (36)..(36)
/ OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (39)..(39)
/ OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (40)..(40)
/ OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
/ NAME/KEY: misc_feature
/ LOCATION: (48)..(48)
/ OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala,
/ NAME/KEY: misc_feature
/ LOCATION: (54)..(54)
/ OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
/ NAME/KEY: misc_feature
/ LOCATION: (56)..(56)
/ OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
/ NAME/KEY: misc_feature
/ LOCATION: (65)..(65)
/ OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
/ NAME/KEY: misc_feature
/ LOCATION: (66)..(66)
/ OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
/ NAME/KEY: misc_feature
/ LOCATION: (76)..(76)
/ OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
/ NAME/KEY: misc_feature
/ LOCATION: (87)..(87)
/ OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (93)..(93)
/ OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
/ NAME/KEY: misc_feature
/ LOCATION: (102)..(102)
/ OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
/ NAME/KEY: misc_feature
/ LOCATION: (103)..(103)
/ OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (109)..(109)
/ OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (112)..(112)
/ OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
/ NAME/KEY: misc_feature
/ LOCATION: (114)..(114)
/ OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (117)..(117)
/ OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
/ NAME/KEY: misc_feature
/ LOCATION: (148)..(148)
/ OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
/ NAME/KEY: misc_feature
/ LOCATION: (150)..(150)
/ OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
/ NAME/KEY: misc_feature
/ LOCATION: (153)..(153)
/ OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (154)..(154)
/ OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
/ NAME/KEY: misc_feature
/ LOCATION: (157)..(157)
/ OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
/ NAME/KEY: misc_feature
/ LOCATION: (183)..(183)
/ OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
/ NAME/KEY: misc_feature
/ LOCATION: (223)..(223)
/ OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
/ NAME/KEY: misc_feature
/ LOCATION: (225)..(225)
/ OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
/ NAME/KEY: misc_feature
/ LOCATION: (237)..(237)
/ OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala,
/ NAME/KEY: misc_feature
/ LOCATION: (239)..(239)
/ OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
/ NAME/KEY: misc_feature
/ LOCATION: (241)..(241)
/ OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
/ NAME/KEY: misc_feature
/ LOCATION: (242)..(242)
/ OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (246)..(246)
/ OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
/ NAME/KEY: misc_feature
/ LOCATION: (251)..(251)
/ OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (259)..(259)
/ OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
/ NAME/KEY: misc_feature
/ LOCATION: (305)..(305)
/ OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
/ NAME/KEY: misc_feature
```

LOCATION: (319)..(319)
OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
NAME/KEY: misc feature
LOCATION: (346)..(346)
OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
OTHER INFORMATION: stop codon, Trp, or Ser.
NAME/KEY: misc feature
LOCATION: (411)..(411)
OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
NAME/KEY: misc feature
LOCATION: (421)..(421)
OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (457)..(457)
OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
NAME/KEY: misc feature
LOCATION: (458)..(458)
OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
NAME/KEY: misc feature
LOCATION: (466)..(466)
OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
NAME/KEY: misc feature
LOCATION: (487)..(487)
OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
NAME/KEY: misc feature
LOCATION: (493)..(493)
OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
NAME/KEY: misc feature
LOCATION: (518)..(518)
OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
NAME/KEY: misc feature
LOCATION: (536)..(536)
OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
NAME/KEY: misc feature
LOCATION: (556)..(556)
OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (557)..(557)
OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (558)..(558)
OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.

Query Match 75.6%; Score 2730; DB 9; Length 720;

Best Local Similarity 76.5%; Pred. No. 2.5e-239;

Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

QY 1 MASPISLSTATTLANGFTNGSHAAPTKSAAGFTSALRTPGDLGHAHQSOLIEYQELLSLD 60
DB 1 MAASXOSIATSGXNGXKXNXXHAAFXKXGTGATSLXRPXKXLAPEPTQYQDIDIXELAD 60
QY 61 PTDDVVELSGSYLTVDVVGAAARKGRVYQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
DB 61 PTDDXXELDGYTLTLDVVGAAARKGRVYQNDDEIRAKIDXXVEFLRXQXKXVYVYVTT 120
QY 121 GFGGSAADRTEDAVSLQKALIEHQLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRVN 180
DB 121 GFGGSAADRTEDAVSLQKALIEHQLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRVN 180
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVLELRSISASGDSLPSLYIGAITGHPDVYH 240
DB 181 SLTRGSAVRLVLEALTNFLNHRITPIVLELRSISASGDSLPSLYIGAITGHPDVYH 240
QY 241 VLHEGTEKIMFAREALISFLGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 241 VLHEGTEKIMFAREALISFLGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLISQ 300
QY 301 ALTALTVEAMVGGGSPFPIHDVCRPHGQVEVARNRITLLSGSFFVHEEVEVKYKD 360
DB 301 ALTALTVEAMVGGGSPFPIHDVCRPHGQVEVARNRITLLSGSFFVHEEVEVKYKD 360
QY 361 EGIILRODRYPLRTSPQFLPVEDWMAVSTLSLNNNTTNDPLIDVENKOTAHGNFOA 420

DB 361 EGIILRODRYPLRTSPQFLPVEDWMAVSTLSLNNNTTNDPLIDVENKOTAHGNFOA 420
QY 421 SAVSISMEXTRLALALIGKLNFTQCTELNNAANRGPLSCLAADPSLHNHGLDHTA 480
DB 421 XAVANTMKRTBLALALIGKLNFTQCTELNNAANRGXSSCLAADPSLHYCGGLDAAA 480
QY 481 AVASELGHANVTTFVQPAEMGQAVNSIALISARRTAEANDVLSLASHLYCTLOAV 540
DB 481 AYTSELHNLANPXTTHVQPREMGQAVNSIALISARRTAEANDVLSLASHLYCTLOAV 540
QY 541 DIRMELDFKQFDBLPPTLLQOHGIGLDVNALALEVKKALNKRLEQTTTYLDEPRWD 600
DB 541 DIRMELDFKQFDBLPPTLLQOHGIGLDVNALALEVKKALNKRLEQTTTYLDEPRWD 600
QY 601 AFSVATGIVVELLSSPSPANTLVANAMKVASAKKASITREVRNRFMOQTPSSQAPHA 660
DB 601 AFSVATGIVVELLSSPSPANTLVANAMKVASAKKASITREVRNRFMOQTPSSQAPHA 660
QY 661 YLSPRTYLYSFVEBELGVQARGDVFGVQOETIGSNVRIYEAIRKGRINHYVWMLA 720
DB 661 YLSPRTYLYSFVEBELGVQARGDVFGVQOETIGSNVRIYEAIRKGRINHYVWMLA 720

RESULT 4

US-09-765-873A-8

Sequence 8, Application US/09765873A

Patent No. US20010053847A1

GENERAL INFORMATION:

APPLICANT: Yang, Xiao-Song

TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID

FILE REFERENCE: BC1009 US CIP

CURRENT APPLICATION NUMBER: US/09/765, 873A

CURRENT FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: US 09/627, 216

PRIOR FILING DATE: 2000-07-27

PRIOR APPLICATION NUMBER: US 60/147, 719

PRIOR FILING DATE: 1999-08-06

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 716

TYPE: PRK

ORGANISM: Rhodotorula glutinis

US-09-765-873A-8

Query Match 72.4%; Score 2614; DB 9; Length 716;

Best Local Similarity 72.0%; Pred. No. 9.5e-229;

Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 1 MASPISLSTATTLANGFTNGSHAAPTKSAAGFTSALRTPGDLGHAHQSOLIEYQELLSLD 60
DB 1 MASPISLSTATTLANGFTNGSHAAPTKSAAGFTSALRTPGDLGHAHQSOLIEYQELLSLD 60
QY 61 PTDDVVELSGSYLTVDVVGAAARKGRVYQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
DB 61 PTDDVVELSGSYLTVDVVGAAARKGRVYQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
QY 121 GFGGSAADRTEDAVSLQKALIEHQLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRVN 180
DB 121 GFGGSAADRTEDAVSLQKALIEHQLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRVN 180
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVLELRSISASGDSLPSLYIGAITGHPDVYH 240
DB 181 SLTRGSAVRLVLEALTNFLNHRITPIVLELRSISASGDSLPSLYIGAITGHPDVYH 240
QY 241 VLHEGTEKIMFAREALISFLGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 241 VLHEGTEKIMFAREALISFLGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLISQ 300
QY 301 ALTALTVEAMVGGGSPFPIHDVCRPHGQVEVARNRITLLSGSFFVHEEVEVKYKD 360
DB 301 ALTALTVEAMVGGGSPFPIHDVCRPHGQVEVARNRITLLSGSFFVHEEVEVKYKD 360
QY 361 EGIILRODRYPLRTSPQFLPVEDWMAVSTLSLNNNTTNDPLIDVENKOTAHGNFOA 420

```

QY      361 EGILRODRYPLRTSPQFLGPIVEDMMAHAYSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 419
      355 EGILRODRYPLRTSPQFLGPIVEDMMAHAYSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 414
      420 ASAVSISMEXTPLALALIGKLNFTQCTELNANMRGLPSCLAEDPSLHNGGLDHI 479
      415 AAIVANMTEKTRIGLAQIGKLNFTQCTELNANMRGLPSCLAEDPSLHNGGLDHI 474
      480 AAIVASLIGHLANPVTTFFVQPAEMGNOAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
      475 AATSTELGHLANPVTTTHVQPAEMANOAVNSLALISARRTTESNDVLSLLATHLYCVLOA 534
      540 VDLRAVELDFKKQDFPLPFTLLQOHLG---TGLDV--NALALEVKKALKRLEQTTYYDLE 595
      535 IDLRAIEFEFFKKQDFGPAIVSLIDQHFSGSAMTGSNLRDELVEKVNKTAKRLEQTNYSYDLV 594
      596 PRMHDAFSYATGTVEVLELSSPSANVTITAVANAMKVASAEKALISTREVRNRFQOTSSQ 655
      595 PRMHDAFSYATGTVEVLELSSPSANVTITAVANAMKVASAEKALISTREVRNRFQOTSSQ 651
      656 APAAHAYLSPTRTVLSFYREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLT 715
      652 SPALSTYLSPTRTQILYAFVREELGVQARRGDVFLGQEVTTIGSNVSKIYEAIKSGRIHVLT 711
      716 VKMLA 720
      712 LKMLA 716

RESULT 5
US-09-939-408a-19
; Sequence 19, Application US/09939408a
; Patent No. US20020102712a1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408a
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula toruloides
US-09-939-408a-19

Query Match      72.4%; Score 2614; DB 9; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1 MAPSLDLSLATTLANGFTNGSHAPTKSAAGPTSAIRRTPGLDGHAHOSQLEIYQELISD 60
      1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTOYDIVERKMLAA 54
      61 PTDVVELSGYSLTVADVGAARKGRVRVONDEIRARVDKSVDFLKAQOUNSYGYGTT 120
      55 PTDSTLELDGYSLNIGDVSAARKGRPVAVKSDIRSKIDKSVDFLSQLSMSYGYGTT 114
      121 GFGGSADRTTEDAVSLQKALIEHOLCGVTPTSXSFSGRGLENTLPLEVVRGAMVIRVN 180
      115 GFGGSADRTTEDAVSLQKALIEHOLCGVTPTSXSFSGRGLENTLPLEVVRGAMVIRVN 174
      181 SLTRGSHAVLVVLELTNTNLRIRPIVPLRGSTISASGDSLPLSYIGATIGHPDVVH 240
      175 SLTRGSHAVLVVLELTNTNLRIRPIVPLRGSTISASGDSLPLSYIAAISGHPDSKVH 234
      Db

```

```

QY      241 VLHEGTEKIMFAREASLIFGLAEVAVLQKKEGLGVNGTAVASAMATLSLHDSAMLISLQ 300
      235 VHEGTEKILYAREAMAFNLLEPVLPGRKEGLGVNGTAVASAMATLSLHDSAMLISLQ 294
      301 AATLTVEAMVQOQSSFPRIHDVCRPHPGQVEYARNRITLLSGSSFAVEEERVKYKD 360
      295 SLTAMTVEAMVGHASFPRLHDVTRPPTQIEVAGNRKLLBSGRPAVHHEERVKYKD 354
      361 EGILRODRYPLRTSPQFLGPIVEDMMAHAYSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 419
      355 EGILRODRYPLRTSPQFLGPIVEDMMAHAYSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 414
      420 ASAVSISMEXTPLALALIGKLNFTQCTELNANMRGLPSCLAEDPSLHNGGLDHI 479
      415 AAIVANMTEKTRIGLAQIGKLNFTQCTELNANMRGLPSCLAEDPSLHNGGLDHI 474
      480 AAIVASLIGHLANPVTTFFVQPAEMGNOAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
      475 AATSTELGHLANPVTTTHVQPAEMANOAVNSLALISARRTTESNDVLSLLATHLYCVLOA 534
      540 VDLRAVELDFKKQDFPLPFTLLQOHLG---TGLDV--NALALEVKKALKRLEQTTYYDLE 595
      535 IDLRAIEFEFFKKQDFGPAIVSLIDQHFSGSAMTGSNLRDELVEKVNKTAKRLEQTNYSYDLV 594
      596 PRMHDAFSYATGTVEVLELSSPSANVTITAVANAMKVASAEKALISTREVRNRFQOTSSQ 655
      595 PRMHDAFSYATGTVEVLELSSPSANVTITAVANAMKVASAEKALISTREVRNRFQOTSSQ 651
      656 APAAHAYLSPTRTVLSFYREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLT 715
      652 SPALSTYLSPTRTQILYAFVREELGVQARRGDVFLGQEVTTIGSNVSKIYEAIKSGRIHVLT 711
      716 VKMLA 720
      712 LKMLA 716

RESULT 6
US-10-138-970A-4
; Sequence 4, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sarsaiani, Sima
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-138-970A-4

Query Match      72.4%; Score 2614; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1 MAPSLDLSLATTLANGFTNGSHAPTKSAAGPTSAIRRTPGLDGHAHOSQLEIYQELISD 60
      1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTOYDIVERKMLAA 54
      61 PTDVVELSGYSLTVADVGAARKGRVRVONDEIRARVDKSVDFLKAQOUNSYGYGTT 120
      55 PTDSTLELDGYSLNIGDVSAARKGRPVAVKSDIRSKIDKSVDFLSQLSMSYGYGTT 114
      121 GFGGSADRTTEDAVSLQKALIEHOLCGVTPTSXSFSGRGLENTLPLEVVRGAMVIRVN 180
      Db

```

Db 115 GFGSSADRTEDAIISLOKALLEHQLCGVLPSSGDSRLGRGLSENSLPLEVVRGANTIVN 174
Qy 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDSLPSLYAGAITGHPDVKA 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGTISASGDSLPSLYAIAAISGHPDKVH 234
Qy 241 VHEGTEKIMFARBAISLFGLEAVVLGPRGELGVNGTVASASMAITLSLHDSHMLSLISQ 300
Db 235 VVHEGKEKILYARBAAMLFNLEPVVLGPKGELGVNGTVASASMAITLALHDAHMLSLISQ 294
Qy 301 ALTALTEAMVGGQSPAPFIHDVCRPHGQVEVARNITRLSGSSFAVEHEEYKVKD 360
Db 295 SLTMTVEAMVGHAGSHRPFILHDVTRPHPTQIEVAGNIRKLLEGRFPAVHEEYKVKD 354
Qy 361 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Db 355 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHKGKLDIHI 479
Db 415 AAAVANTMEKTRIGLAIQIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHKGKLDIHI 474
Qy 480 AAAYSELGHLANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSILLASHLYCTLOA 539
Db 475 AAAYSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSILLASHLYCTLOA 534
Qy 540 VDLRAMELDPKQOPDELLPTLLOQHIG---TGLDV-NALALEYKALKRELEQTTYYDLE 595
Db 535 IDLRALIEEFKQFQPAIVSLIDQHFSGAMTGSNLRDELVEKNVTLAKRELEQTNYSYDLV 594
Qy 596 PRMHDAFSYATGTVEELSSPSANVTLTAVNAKVASAKAISLTREVRNRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEELSSPSANVTLTAVNAKVASAKAISLTREVRNRFWQTPSSQ 651
Qy 656 APAAAYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKQGRINHL 715
Db 652 SPALSYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKQGRINHL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 7
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; PRIORITY FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula glutinis
US-10-188-523B-8

Query Match 72.4%; Score 2614; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 9,5e-223;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
Qy 1 MAPSLDSTATTLANGFTNGSHAAPTKSAGPTSALETRTPGLDGHAAHQSCLEIVQELSLD 60
Db 1 MAPSLDSTATSHSPANG-----VASAKQAVNGASTYLAVAGSHLPTTQVTVQVDIYKMLAA 54

Qy 61 PTDDVLEISGYSILTVRDVYGAARKGRVRYONDEIRARDKSVDFKAQLONSVYGVTT 120
Db 55 PFDSTLELDGYSILNTGDVYSAARKGRPVRYADSDPEIRKIDKXVEFRLSQLSMNVYGVTT 114
Qy 121 GFGSSADRTEDAVLSLOKALLEHQLCGVLPSSGDSRLGRGLSENSLPLEVVRGANTIVN 180
Db 115 GFGSSADRTEDAVLSLOKALLEHQLCGVLPSSGDSRLGRGLSENSLPLEVVRGANTIVN 174
Qy 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDSLPSLYAGAITGHPDVKA 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGTISASGDSLPSLYAIAAISGHPDKVH 234
Qy 241 VHEGTEKIMFARBAISLFGLEAVVLGPRGELGVNGTVASASMAITLSLHDSHMLSLISQ 300
Db 235 VVHEGKEKILYARBAAMLFNLEPVVLGPKGELGVNGTVASASMAITLALHDAHMLSLISQ 294
Qy 301 ALTALTEAMVGGQSPAPFIHDVCRPHGQVEVARNITRLSGSSFAVEHEEYKVKD 360
Db 295 SLTMTVEAMVGHAGSHRPFILHDVTRPHPTQIEVAGNIRKLLEGRFPAVHEEYKVKD 354
Qy 361 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Db 355 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHKGKLDIHI 479
Db 415 AAAVANTMEKTRIGLAIQIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHKGKLDIHI 474
Qy 480 AAAYSELGHLANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSILLASHLYCTLOA 539
Db 475 AAAYSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSILLASHLYCTLOA 534
Qy 540 VDLRAMELDPKQOPDELLPTLLOQHIG---TGLDV-NALALEYKALKRELEQTTYYDLE 595
Db 535 IDLRALIEEFKQFQPAIVSLIDQHFSGAMTGSNLRDELVEKNVTLAKRELEQTNYSYDLV 594
Qy 596 PRMHDAFSYATGTVEELSSPSANVTLTAVNAKVASAKAISLTREVRNRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEELSSPSANVTLTAVNAKVASAKAISLTREVRNRFWQTPSSQ 651
Qy 656 APAAAYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKQGRINHL 715
Db 652 SPALSYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKQGRINHL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 8
US-10-374-366-2
; Sequence 2, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; PRIORITY FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,219
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-2

Query Match 72.4%; Score 2614; DB 15; Length 716;
Best Local Similarity 72.0%; Pred. No. 9,5e-223;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

PRIOR APPLICATION NUMBER: US 09/627,216
 PRIOR FILING DATE: 2000-07-27
 PRIOR APPLICATION NUMBER: US 60/147,719
 PRIOR FILING DATE: 1999-08-06
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 35
 LENGTH: 716
 TYPE: PRT
 ORGANISM: mutant from Rhodotorula glutinis
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: (198)..(198)
 OTHER INFORMATION: X= Asp, Asn, Glu, Gln
 NAME/KEY: UNSURE
 LOCATION: (540)..(540)
 OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
 US-09-765-873a-35

Query Match 72.4%; Score 2613; DB 9; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.2e-228;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSLATTLANGFTNGSHAPTKSAAGPTSAURTPGLDGHAAHQSLEIYQELISD 60
 DB 1 MAPSLDLSHSFANG-----VASAKQAVNGASTMLAVAGSHLPTQVTVDIVEXMLAA 54
 QY 61 PTDVVELSGYSTIVDVYGAARKGRVAVQNDDEIRARVDSVDFLAQIQNSVYGVTT 120
 DB 55 PTSTLIEDGYSINLDGVSAARKGRVAVQNDDEIRARVDSVDFLAQIQNSVYGVTT 114
 QY 121 GFGSADTRTEDAVSLQKALIEHQLCVPTXSFSVQGLNTPLEVRGAMVIRVN 180
 DB 115 GFGSADTRTEDAVSLQKALIEHQLCVPTXSFSVQGLNTPLEVRGAMVIRVN 174
 QY 181 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPISYAGAITGHPDKVH 240
 DB 175 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPISYAGAITGHPDKVH 234
 QY 241 VLHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
 DB 235 VVHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
 QY 301 ALTALTVEAMVGGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
 DB 295 SLTAMTVEAMVGGHGSFHPFLHDVTRPHPTQIEVAGNIRKLLGSRFAVHHEEVKXKD 354
 QY 361 EGLRQDRYPLRTSPQFLGPIVEDMMHAYSTLSL-NTTTDNPILDVENKQTAHGNFQ 419
 DB 355 EGLRQDRYPLRTSPQFLGPIVEDMMHAYSTLSL-NTTTDNPILDVENKQTAHGNFQ 414
 QY 420 ASAVSISMEKRLALALIGKLNFTQCTELLNANMNGLPSCLAADPSPINHYGKGLDHI 479
 DB 415 AAAVANTMEKRLALALIGKLNFTQCTELLNANMNGLPSCLAADPSPINHYGKGLDHI 474
 QY 480 AAASSELGHLANPVTTFVQPAEMGNQAVNSLALISARTTAANDVLSLLASHLYCTLOA 539
 DB 475 AAATSELGHLANPVTTFVQPAEMGNQAVNSLALISARTTAANDVLSLLASHLYCTLOA 534
 QY 540 VDLRAVELDFKKQDFLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTVDLE 595
 DB 535 IDIRAKYEFKQDFGALVSLIDQHFSGAMGNSNRDELVERKAKTLAKLEQNSYDLY 594
 QY 596 PRNHDAFSYATGTVELLSSSPSNAVTLTAVANMVASAEKALSLTEVRNRFWQTPSSQ 655
 DB 595 PRNHDAFSYATGTVELLSSSPSNAVTLTAVANMVASAEKALSLTEVRNRFWQTPSSQ 651
 QY 656 APAAHATLSPRTVLVSFVVEELGVQARGDVVVGQOCTIGSNSTRILEYALIKORINHL 715
 DB 652 SPALSTLSPRTVLVSFVVEELGVQARGDVVVGQOCTIGSNSTRILEYALIKORINHL 711
 QY 716 VKMLA 720
 :|||

DB 712 LKMLA 716

RESULT 11
 US-10-138-970A-19
 Sequence 19, Application US/10138970A
 Publication No. US20030079255A1
 GENERAL INFORMATION:
 APPLICANT: Sarsilani, Sima
 APPLICANT: Tang, Xiao-Song
 APPLICANT: Qi, Wei Wei
 TITLE OF INVENTION: Methods for the production of tyrosine, cinnamic acid and para-
 FILE REFERENCE: C11777
 CURRENT APPLICATION NUMBER: US/10/138, 970A
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 19
 LENGTH: 716
 TYPE: PRT
 ORGANISM: Rhodotorula glutinis
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (198)..(198)
 OTHER INFORMATION: X=Asp, Asn, Glu, Gln
 NAME/KEY: MISC FEATURE
 LOCATION: (540)..(540)
 OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
 US-10-138-970A-19

Query Match 72.4%; Score 2613; DB 14; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.2e-228;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSLATTLANGFTNGSHAPTKSAAGPTSAURTPGLDGHAAHQSLEIYQELISD 60
 DB 1 MAPSLDLSHSFANG-----VASAKQAVNGASTMLAVAGSHLPTQVTVDIVEXMLAA 54
 QY 61 PTDVVELSGYSTIVDVYGAARKGRVAVQNDDEIRARVDSVDFLAQIQNSVYGVTT 120
 DB 55 PTSTLIEDGYSINLDGVSAARKGRVAVQNDDEIRARVDSVDFLAQIQNSVYGVTT 114
 QY 121 GFGSADTRTEDAVSLQKALIEHQLCVPTXSFSVQGLNTPLEVRGAMVIRVN 180
 DB 115 GFGSADTRTEDAVSLQKALIEHQLCVPTXSFSVQGLNTPLEVRGAMVIRVN 174
 QY 181 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPISYAGAITGHPDKVH 240
 DB 175 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPISYAGAITGHPDKVH 234
 QY 241 VLHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
 DB 235 VVHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
 QY 301 ALTALTVEAMVGGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
 DB 295 SLTAMTVEAMVGGHGSFHPFLHDVTRPHPTQIEVAGNIRKLLGSRFAVHHEEVKXKD 354
 QY 361 EGLRQDRYPLRTSPQFLGPIVEDMMHAYSTLSL-NTTTDNPILDVENKQTAHGNFQ 419
 DB 355 EGLRQDRYPLRTSPQFLGPIVEDMMHAYSTLSL-NTTTDNPILDVENKQTAHGNFQ 414
 QY 420 ASAVSISMEKRLALALIGKLNFTQCTELLNANMNGLPSCLAADPSPINHYGKGLDHI 479
 DB 415 AAAVANTMEKRLALALIGKLNFTQCTELLNANMNGLPSCLAADPSPINHYGKGLDHI 474
 QY 480 AAASSELGHLANPVTTFVQPAEMGNQAVNSLALISARTTAANDVLSLLASHLYCTLOA 539

```

Db 475 AAYTSELGLHNPVTHQPAEMANQAVNSIALISARTTESNDVLSILLATLHYCYLOA 534
QY 540 VDLRMETLDFKKQDFPLPTLLQOHLG---TGLDV-NALALEYKALKNELEQTTTYDLE 595
Db 535 IDLRAXEPEFKQGFPAIVSLIDQFSGAMTGSNLRBELVEKNKTLAKLEQTNSTYDLV 594
QY 596 PRMHDAFSYATGTVEELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 651
QY 656 APAHAYLSPRTVLYSPFREELGYQARGDVYVGVQOETIGSNRSRYEAIKQGRINHVL 715
Db 652 SPALSYLSPTQILYAFVREELGVARRGDVFLKQEVTTIGSNVSKTYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 12
US-10-188-523B-35
; Sequence 35, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; PRIOR FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis mutant
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-188-523B-35

Query Match 72.4%; Score 2613; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 1,2e-228;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

```

QY 301 ALTAITVAMVGOQGSFAPFTHVCRPHQGVARNIRITLLSGSSFAVEEBEVKVKD 360
Db 295 SLTAMTAVAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLKSGSRPAVHHEBEVKXD 354
QY 361 EGLRQDRYPTRTSPQPLGLPYEDMMTAYSTLSF-NNTTNDPILDVENEKQAHGNGFQ 419
Db 355 EGLRQDRYPTRTSPQPLGLPYEDMMTAYSTLSF-NNTTNDPILDVENEKQAHGNGFQ 414
QY 420 ASAVISMEKTRMLALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNHKGGLDIHI 479
Db 415 AAANVTNEKTRGLGAQKGNFTQLEMTNAGNRRGLPSCLAEDPSLNHKGGLDIHI 474
QY 480 AAYASELGHANPVTTVPQPAEMNQAVNSIALISARTTEANDVLSILLASHLYCYLOA 539
Db 475 AAYTSELGLHNPVTHQPAEMANQAVNSIALISARTTESNDVLSILLATLHYCYLOA 534
QY 540 VDLRMETLDFKKQDFPLPTLLQOHLG---TGLDV-NALALEYKALKNELEQTTTYDLE 595
Db 535 IDLRAXEPEFKQGFPAIVSLIDQFSGAMTGSNLRBELVEKNKTLAKLEQTNSTYDLV 594
QY 596 PRMHDAFSYATGTVEELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 655
Db 595 PRMHDAFSYATGTVEELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 651
QY 656 APAHAYLSPRTVLYSPFREELGYQARGDVYVGVQOETIGSNRSRYEAIKQGRINHVL 715
Db 652 SPALSYLSPTQILYAFVREELGVARRGDVFLKQEVTTIGSNVSKTYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

```

```

RESULT 13
US-10-439-478-15
; Sequence 15, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariassani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Vaneijl, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912
; CURRENT APPLICATION NUMBER: US/10/439,478
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 716
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-439-478-15

```

```

Query Match 72.4%; Score 2613; DB 15; Length 716;
Best Local Similarity 72.0%; Pred. No. 1,2e-228;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

```

QY 1 MAPSLDLSLATTIANGFTNGSHAAPTKSAAGPTSAIRRTGCLDGHAAHQSLIYVOELLSD 60
DB 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDVVELSGYSLTRVDVGAARKGRVRVQNDDEIRARVKSVDPLKAOQNSVYGVT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRVPRVKDSDEIRKSKIDKSVETLRQSLSMSYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 180
DB 115 GFGGSADRTEDASLQKALLEHQLCGYLPSSFSFRLGRGLENSLPLEVVGAMTIRVN 174
QY 181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGISASGDSLPLSTAGATIGHDPVYH 240
DB 175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGISASGDSLPLSTAGATIGHDPVYH 234
QY 241 VLHGTGKIMFARBAISLFGLEAVVLGPKEGLGLVNGTAVASASMATLSLHSHMLSLSQ 300
DB 235 VLHGTGKEXILYARBAAMALFPLEPVLPKEGLGLVNGTAVASASMATLSLHSHMLSLSQ 294
QY 301 ALTLTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNIRTLSSGSFAVHEEVEVKYKD 360
DB 295 SLTMTVEAMVGAHGSFHPFLHDVTRPHPTQIEVAGNIRKLEGRFAVHEEVEVKYKD 354
QY 361 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHGNFQ 419
DB 355 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISEKTRIALALIGKLTNCTQTELLNAAMNRLPSCLAEDPSILNHGKGLDIH 479
DB 415 AAIVANTMEKTRIGLQIGKLTNCTQTELLNAAMNRLPSCLAEDPSILNHGKGLDIH 474
QY 480 AAIVASELGHANPVTTFVOPAEKMGNOAVNSLALISARTEANDVSLILASHLYCTLOA 539
DB 475 AAIVASELGHANPVTTFVOPAEKMGNOAVNSLALISARTEANDVSLILASHLYCTLOA 534
QY 540 VDIRAMELDFKQKPEDPLPTLLQOHLG---TGLDV-NALALEVKAALKRLEQTTVDLE 595
DB 535 IDIRAEFEFKQKQFQGAIVSLIDQFSGAMTGSNLRDELVEKNKTLAKRLEQNTSDLV 594
QY 596 PRMHDAFSYATGTVEVLLSSPSBANTLTAVNAMKVASAKAISLITREVRNRWQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLLSSPSBANTLTAVNAMKVASAKAISLITREVRNRWQTPSSQ 651
QY 656 APAHAYLSPTRTVLSFVREELGVQARGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSYLSPTRTQILYAFVREELGVQARGDVFLGQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

```

RESULT 14

```

/ Sequence 10, Application US/09765873A
/ Patent No. US20010053847A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/ FILE REFERENCE: BC1009 US CIP
/ CURRENT APPLICATION NUMBER: US/09/765, 873A
/ PRIOR FILING DATE: 2001-01-19
/ PRIOR APPLICATION NUMBER: US 09/627, 216
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR APPLICATION NUMBER: US 60/147, 719
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodotorula glutinis

```

US-09-765-873A-10

Query Match 72.4%; Score 2612; DB 9; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1,4e-228;
 Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

QY 1 MAPSLDLSLATTIANGFTNGSHAAPTKSAAGPTSAIRRTGCLDGHAAHQSLIYVOELLSD 60
DB 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDVVELSGYSLTRVDVGAARKGRVRVQNDDEIRARVKSVDPLKAOQNSVYGVT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRVPRVKDSDEIRKSKIDKSVETLRQSLSMSYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 180
DB 115 GFGGSADRTEDASLQKALLEHQLCGYLPSSFSFRLGRGLENSLPLEVVGAMTIRVN 174
QY 181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGISASGDSLPLSTAGATIGHDPVYH 240
DB 175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGISASGDSLPLSTAGATIGHDPVYH 234
QY 241 VLHGTGKIMFARBAISLFGLEAVVLGPKEGLGLVNGTAVASASMATLSLHSHMLSLSQ 300
DB 235 VLHGTGKEXILYARBAAMALFPLEPVLPKEGLGLVNGTAVASASMATLSLHSHMLSLSQ 294
QY 301 ALTLTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNIRTLSSGSFAVHEEVEVKYKD 360
DB 295 SLTMTVEAMVGAHGSFHPFLHDVTRPHPTQIEVAGNIRKLEGRFAVHEEVEVKYKD 354
QY 361 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHGNFQ 419
DB 355 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISEKTRIALALIGKLTNCTQTELLNAAMNRLPSCLAEDPSILNHGKGLDIH 479
DB 415 AAIVANTMEKTRIGLQIGKLTNCTQTELLNAAMNRLPSCLAEDPSILNHGKGLDIH 474
QY 480 AAIVASELGHANPVTTFVOPAEKMGNOAVNSLALISARTEANDVSLILASHLYCTLOA 539
DB 475 AAIVASELGHANPVTTFVOPAEKMGNOAVNSLALISARTEANDVSLILASHLYCTLOA 534
QY 540 VDIRAMELDFKQKPEDPLPTLLQOHLG---TGLDV-NALALEVKAALKRLEQTTVDLE 595
DB 535 IDIRAEFEFKQKQFQGAIVSLIDQFSGAMTGSNLRDELVEKNKTLAKRLEQNTSDLV 594
QY 596 PRMHDAFSYATGTVEVLLSSPSBANTLTAVNAMKVASAKAISLITREVRNRWQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLLSSPSBANTLTAVNAMKVASAKAISLITREVRNRWQTPSSQ 651
QY 656 APAHAYLSPTRTVLSFVREELGVQARGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSYLSPTRTQILYAFVREELGVQARGDVFLGQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

```

RESULT 15

```

/ Sequence 24, Application US/10138970A
/ Publication No. US20030079255A1
/ GENERAL INFORMATION:
/ APPLICANT: Sariastani, Sima
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei-Mei
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ FILE REFERENCE: C11777
/ CURRENT APPLICATION NUMBER: US/10/138, 970A
/ PRIOR FILING DATE: 2002-07-23
/ NUMBER OF SEQ ID NOS: 24

```



```
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO: 24
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24
```

Query Match 72.4%; Score 2612; DB 14; Length 716;

Best Local Similarity 72.0%; Pred. No. 1,4e-228; Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```
QY 1 MAPSLDLSLATTANGFTNGSHAPTKSAAGPTSAALRRTPGLDGHAHQSOLEIVQELSD 60
DB 1 MAPSLDSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVYDIKMLAA 54
QY 61 PTDVVELSGYGLTVRDVVGAAKRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTDSTLELDGSLNIDVVASARKGRPVYKDSDEIRSKIDKSVDFLRSQLSMSVYGVTT 114
QY 121 GFGGASDTRTEDAVSLQKALIEHOLCGVPTSSSFSVGRGIENTLPLEVVRGAMVIRVN 180
DB 115 GFGGASDTRTEDAISIQKALLEHQLCGVLPSSFDSFRLGRGLNSLPLEVRGAMTIRVN 174
QY 181 SLTRGSAVRLVLEALTNFNHRTTPVPLRGSISASGDSPLSYIAGAITGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFNHRTTPVPLRGTISASGDSPLSYIAAISGHPDSKH 234
QY 241 VHEGTEKIMFARALSLFGLAAYVLGREGGLVNGTAVASAMATLSLHSHMLSLSQ 300
DB 235 VHEGEEKILYAREKAPALFNLEPVVLGPEEGLVNGTAVASAMATLALHDAHMLSLSQ 294
QY 301 ALTALVTEAMVQOQSFAFFIDVCRPHPGQVEVARNIRTLSSGSPAVEHEEEVKYKD 360
DB 295 SLTAMVTEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGRFVHHHEEVKXKD 354
QY 361 EGLRQDRYPLRTSPQFLGPIVEDMMAHAYSTLSLE-NNTTDNPLLDVENKQTAHGNFQ 419
DB 355 EGLRQDRYPLRTSPQFLGPIVSLDIHAHAVALTIKAGOSTDNDPLIDVENKTSHHGNFQ 414
QY 420 ASAVSISMEKTLALALGKLNFTQCTELNAAMNRGLPSCIAAEDPSLNYHGKLDIHI 479
DB 415 AAANVTMEKTRLGLAQIGKLNFTQCTELNAGMRGLPSCIAAEDPSLNYHGKLDIAA 474
QY 480 AAYASELGHANPVTTFFVQPAEMGNQAVNSLALISARRTABANDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHANPVTTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
QY 540 VDIRAMELDFKQFDPDLPTLLOQHGL--TGLDV-NALALEYKALKRLLEQTTTYDLE 595
DB 535 IDLRATEFEFKQFGPAIVSLIDHFGSAMTGSNLDLVEKVKTKLAKRLEQTNSTYDLV 594
QY 596 PRWHDASFATGIVVELSSPSANVTLLAVAMKVASAEKASITREVRNRFWQTPSSQ 655
DB 595 PRWHDASFATGIVVELSSPSANVTLLAVAMKVASAEKASITREVRNRFWQTPSSQ 651
QY 656 APAAHAYLSPTRYLSFVEEELGVQARRQDVFGVQOETIGSNVRIYEAIKQGRINHL 715
DB 652 SPALSTYLSPTQIILYAFVEEELGVKARRQDVFLGKQEVYIGSNVSKIYEAIKSGRINVL 711
QY 716 VKMLA 720
DB 712 LKMLA 716
```

Search completed: September 9, 2004, 10:24:24
Job time: 95.6141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:04:14 ; Search time 295.768 Seconds

(without alignments)
2376.050 Million cell updates/sec

Title: US-09-939-408a-13

Perfect score: 3610
Sequence: 1 MAPSIDSATLTIANGFTNGS.....RIYRAIDGRINHYLMQLA 720

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 segs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:*

1:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
2:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22:	/cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
23:	/cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24:	/cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25:	/cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26:	/cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27:	/cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28:	/cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29:	/cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30:	/cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31:	/cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32:	/cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33:	/cgn2_6/ptodata/2/paa/US00_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	3608	99.9	720	1	PCT-US02-26934A-13	Sequence 13, Appl
2	3608	99.9	720	24	US-09-939-408A-13	Sequence 13, Appl
3	2770	76.7	726	1	PCT-US02-26934A-21	Sequence 21, Appl
4	2770	76.7	726	24	US-09-939-408A-21	Sequence 21, Appl
5	2730	75.6	720	1	PCT-US02-26934A-30	Sequence 30, Appl
6	2730	75.6	720	24	US-09-939-408A-30	Sequence 30, Appl
7	2614	72.4	716	1	PCT-US02-26934A-19	Sequence 19, Appl
8	2614	72.4	716	1	PCT-US03-05708-2	Sequence 2, Appl
9	2614	72.4	716	1	PCT-US03-17926-2	Sequence 2, Appl
10	2614	72.4	716	24	US-09-939-408A-19	Sequence 19, Appl
11	2614	72.4	716	27	US-10-138-970A-19	Sequence 8, Appl
12	2614	72.4	716	27	US-10-138-970A-8	Sequence 8, Appl
13	2614	72.4	716	27	US-10-188-523C-8	Sequence 2, Appl
14	2614	72.4	716	29	US-10-374-366-2	Sequence 2, Appl
15	2614	72.4	716	33	US-60-439-478-2	Sequence 2, Appl
16	2614	72.4	716	33	US-60-147-719-8	Sequence 8, Appl
17	2614	72.4	716	33	US-60-383-089-4	Sequence 8, Appl
18	2614	72.4	716	33	US-60-460-120-4	Sequence 4, Appl
19	2613	72.4	716	1	PCT-US03-17926-15	Sequence 15, Appl
20	2613	72.4	716	27	US-10-138-970A-19	Sequence 19, Appl
21	2613	72.4	716	27	US-10-188-523B-19	Sequence 35, Appl
22	2613	72.4	716	27	US-10-188-523C-35	Sequence 35, Appl
23	2613	72.4	716	30	US-10-439-478-15	Sequence 15, Appl
24	2612	72.4	716	1	PCT-US03-05708-4	Sequence 4, Appl
25	2612	72.4	716	27	US-10-138-970A-24	Sequence 24, Appl
26	2612	72.4	716	27	US-10-188-523B-10	Sequence 10, Appl
27	2612	72.4	716	27	US-10-188-523C-10	Sequence 10, Appl
28	2612	72.4	716	29	US-60-147-719-10	Sequence 4, Appl
29	2612	72.4	716	33	US-60-374-366-4	Sequence 10, Appl
30	2609	72.3	716	1	PCT-US03-05708-24	Sequence 24, Appl
31	2609	72.3	716	29	US-10-374-366-24	Sequence 24, Appl
32	2608	72.2	716	1	PCT-US03-05708-22	Sequence 22, Appl
33	2608	72.2	716	29	US-10-374-366-22	Sequence 22, Appl
34	2607	72.2	716	1	PCT-US03-05708-16	Sequence 16, Appl
35	2607	72.2	716	1	PCT-US03-17926-16	Sequence 16, Appl
36	2607	72.2	716	27	US-10-138-970A-22	Sequence 22, Appl
37	2607	72.2	716	27	US-10-188-523B-38	Sequence 38, Appl
38	2607	72.2	716	27	US-10-188-523C-38	Sequence 38, Appl
39	2607	72.2	716	29	US-10-374-366-16	Sequence 16, Appl
40	2607	72.2	716	30	US-10-439-478-18	Sequence 18, Appl
41	2606	72.2	716	1	PCT-US03-05708-8	Sequence 8, Appl
42	2606	72.2	716	29	US-10-374-366-8	Sequence 10, Appl
43	2605	72.2	716	1	PCT-US03-05708-10	Sequence 10, Appl
44	2605	72.2	716	29	US-10-374-366-10	Sequence 10, Appl
45	2604	72.1	716	1	PCT-US03-05708-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
PCT-US02-26934A-13
Sequence 13, Application PC/TUS0226934A
GENERAL INFORMATION:
APPLICANT: PDBU Services, Inc. et al.
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
TITLE OF INVENTION: Using Same
FILE REFERENCE: 29479/500SCA PCT
CURRENT APPLICATION NUMBER: PCT/US02/26934A
CURRENT FILING DATE: 2002-08-23
PRIOR APPLICATION NUMBER: US 09/939,408
PRIOR FILING DATE: 2001-08-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodotorula graminis
FEATURE:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala

Mon Sep 13 10:31:10 2004

us-09-939-408a-13.rapm

Page 2

PCT-US02-26934A-13

Query Match 99.9%; Score 3608; DB 1; Length 720;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIYOELLSD 60
D 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIYOELLSD 60
QY 61 PTDDVVELSGYSLTVRDVVGAAARKGRVYQNDDEIRAVDKSVDFLKAQLONSVYGVTT 120
D 61 PTDDVVELSGYSLTVRDVVGAAARKGRVYQNDDEIRAVDKSVDFLKAQLONSVYGVTT 120
QY 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
D 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
QY 181 SLTRGSAVRVLYVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGAITGHPDKVH 240
D 181 SLTRGSAVRVLYVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGAITGHPDKVH 240
QY 241 VHEGTEKIMFAREASLFGLEAVVLPKREGLVNGTAVASAMATLSLHSHMLSLLSQ 300
D 241 VHEGTEKIMFAREASLFGLEAVVLPKREGLVNGTAVASAMATLSLHSHMLSLLSQ 300
QY 301 ALTALTEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRLLSGSSFAVEHEEVKXKD 360
D 301 ALTALTEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRLLSGSSFAVEHEEVKXKD 360
QY 361 EGIIRDYRPLRTSPQFPLVEDMMHAYSTLSLNNTTTNDPLDVENKOTAHGNFOA 420
D 361 EGIIRDYRPLRTSPQFPLVEDMMHAYSTLSLNNTTTNDPLDVENKOTAHGNFOA 420
QY 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCLAEDPSLNYHKGDLIHIA 480
D 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCLAEDPSLNYHKGDLIHIA 480
QY 481 AVASELGHLANPVTTPVQPAEMGNQAVNSLISARRTAANDVLSLLASHLYCTLOAV 540
D 481 AVASELGHLANPVTTPVQPAEMGNQAVNSLISARRTAANDVLSLLASHLYCTLOAV 540
QY 541 DLRAMELDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWD 600
D 541 DLRAMELDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWD 600
QY 601 AFSYATGTVELLSSSSANVTLLAVNAKVASAKAISLITREVRNRFPQTPSSQAPAH 660
D 601 AFSYATGTVELLSSSSANVTLLAVNAKVASAKAISLITREVRNRFPQTPSSQAPAH 660
QY 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLYKMLA 720
D 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLYKMLA 720
```

RESULT 2
US-09-939-408a-13
Sequence 13, Application US/09939408A
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939,408A
PRIOR FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodotorula graminis
FEATURE:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408a-13

Query Match 99.9%; Score 3608; DB 24; Length 720;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIYOELLSD 60
D 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIYOELLSD 60
QY 61 PTDDVVELSGYSLTVRDVVGAAARKGRVYQNDDEIRAVDKSVDFLKAQLONSVYGVTT 120
D 61 PTDDVVELSGYSLTVRDVVGAAARKGRVYQNDDEIRAVDKSVDFLKAQLONSVYGVTT 120
QY 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
D 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
QY 181 SLTRGSAVRVLYVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGAITGHPDKVH 240
D 181 SLTRGSAVRVLYVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGAITGHPDKVH 240
QY 241 VHEGTEKIMFAREASLFGLEAVVLPKREGLVNGTAVASAMATLSLHSHMLSLLSQ 300
D 241 VHEGTEKIMFAREASLFGLEAVVLPKREGLVNGTAVASAMATLSLHSHMLSLLSQ 300
QY 301 ALTALTEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRLLSGSSFAVEHEEVKXKD 360
D 301 ALTALTEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRLLSGSSFAVEHEEVKXKD 360
QY 361 EGIIRDYRPLRTSPQFPLVEDMMHAYSTLSLNNTTTNDPLDVENKOTAHGNFOA 420
D 361 EGIIRDYRPLRTSPQFPLVEDMMHAYSTLSLNNTTTNDPLDVENKOTAHGNFOA 420
QY 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCLAEDPSLNYHKGDLIHIA 480
D 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCLAEDPSLNYHKGDLIHIA 480
QY 481 AVASELGHLANPVTTPVQPAEMGNQAVNSLISARRTAANDVLSLLASHLYCTLOAV 540
D 481 AVASELGHLANPVTTPVQPAEMGNQAVNSLISARRTAANDVLSLLASHLYCTLOAV 540
QY 541 DLRAMELDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWD 600
D 541 DLRAMELDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWD 600
QY 601 AFSYATGTVELLSSSSANVTLLAVNAKVASAKAISLITREVRNRFPQTPSSQAPAH 660
D 601 AFSYATGTVELLSSSSANVTLLAVNAKVASAKAISLITREVRNRFPQTPSSQAPAH 660
QY 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLYKMLA 720
D 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVLYKMLA 720
```

RESULT 3
PCT-US02-26934A-21
Sequence 21, Application PC/TUS0226934A
GENERAL INFORMATION:
APPLICANT: PCB Services, Inc. et al.
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA PCT
CURRENT APPLICATION NUMBER: PCT/US02/26934A

```
/ CURRENT FILING DATE: 2002-08-23
/ PRIOR APPLICATION NUMBER: US 09/939,408
/ PRIOR FILING DATE: 2001-08-24
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 726
/ TYPE: PR
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (12)..(719)
/ OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that positio
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Consensus of
/ OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
PCT-US02-26934A-21

Query Match      76.7%; Score 2770; DB 1; Length 726;
Best Local Similarity 77.7%; Pred. No. 4.3e-262;
Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

Db      1 MAPSDSLATTLANGFTNGSHAPTKSAAGPTSAALRRTPGLDGHAAHQSLIYQELISD 60
1 MAPSDSLATTSXANGXNKGHAAXXASXXXXXXASXSLPTTXXTQLDIYEXXSLAD 60

QY      61 P-TDVEISGYSITVRDVGAAKGRVAVQNDDEIRAVYDKSVDFLKAQLQNSYGV 119
61 PXTDXXXELDGSYSLTGIVGAAKGRVAVYXDSDEIRXKIDKSVDFLRXQLXNSYGV 120

Db      120 TGFSGSADRTEDAVSLQKALIEHQLGVPPTSXSFVSGRGENTLPLEVRGAMVIRV 179
121 TGFSGSADRTEDAVSLQKALIEHQLGVPPTSXSFVSGRGENTLPLEVRGAMVIRV 180

QY      180 NSLTRGHSANRLVLEALTNLNHRITPIVPLRGISASAGDLSPLSYIAGAITGHPDV 239
181 NSLTRGHSANRLVLEALTNLNHRITPIVPLRGISASAGDLSPLSYIAAITGHPDSKV 240

Db      240 HYLHGETKTFAPARAISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHDSHMLSL 239
241 HXHEGKEXIKKARAIALFGLPEVVLGPKGGLGVNGTAVASAMATLALHDAHMLSL 300

QY      300 QALTALTYEAMVGQGSFAPFIHDVCRPHQGVAVANIRITLSSGSAFVHEHEEVK 359
301 QALTALTYEAMVGAGSHHPFLHDVTRHPQGLEAVANIRITLSSGSAFVHEHEEVK 360

Db      360 DEGIIRQDRYPLRTSPQPLGVLVEDMMAVSTLSL-NTTTDPLLDVENKQTAHGNF 418
361 DEGIIRQDRYPLRTSPQPLGVLVEDMMAVSTLSL-NTTTDPLLDVENKQTAHGNF 420

QY      419 QASAVISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 478
421 QASAVXNTEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 480

Db      479 IAAVASELGHLANPVTTFVQPAEMNGQAVNSIALISARTEANDVLSLLASHLYCTIQ 538
481 AAAYTSELGHLANPVTTFVQPAEMNGQAVNSIALISARTEANDVLSLLASHLYCTIQ 540

QY      539 AVDLRAMELDFKKQDPLPTLLOQHJGTGLDVNA---LALVYKXKLNKRLLEQTTTYDL 594
541 AVDLRAMELDFKKQDPLPTLLOQHJGTGLDVNA---LALVYKXKLNKRLLEQTTTYDL 600

Db      595 EPRMHDFAFYANGTVVELISSPSANVTLTAVNAKVASAEKALSTREYANRFWQFSS 654
601 EPRMHDFAFYANGTVVELISSPSANVTLTAVNAKVASAEKALSTREYANRFWQFSS 660

QY      655 QAPAAHAYSPRRVLYSPVREELGVAQARGDVFGVQOETIGSNVSRITYEAIKGRINHV 714
661 SSPALAYLSPRRVLYSPVREELGVAQARGDVFGVQOETIGSNVSRITYEAIKGRINHV 720

Db      715 LVKMLA 720
721 LVKMLA 726
```

```
RESULT 4
US-09-939-408a-21
/ Sequence 21, Application US/09939408a
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ FILE REFERENCE: 29479/500NSCA
/ CURRENT APPLICATION NUMBER: US/09/939,408A
/ CURRENT FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 09/624,693
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ PRIOR FILING DATE: 2001-07-24
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 726
/ TYPE: PR
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (12)..(719)
/ OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that
/ OTHER INFORMATION: Description of Artificial Sequence: Consensus of
US-09-939-408a-21

Query Match      76.7%; Score 2770; DB 24; Length 726;
Best Local Similarity 77.7%; Pred. No. 4.3e-262;
Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

QY      1 MAPSDSLATTLANGFTNGSHAPTKSAAGPTSAALRRTPGLDGHAAHQSLIYQELISD 60
1 MAPSDSLATTSXANGXNKGHAAXXASXXXXXXASXSLPTTXXTQLDIYEXXSLAD 60

Db      61 P-TDVEISGYSITVRDVGAAKGRVAVQNDDEIRAVYDKSVDFLKAQLQNSYGV 119
61 PXTDXXXELDGSYSLTGIVGAAKGRVAVYXDSDEIRXKIDKSVDFLRXQLXNSYGV 120

QY      120 TGFSGSADRTEDAVSLQKALIEHQLGVPPTSXSFVSGRGENTLPLEVRGAMVIRV 179
121 TGFSGSADRTEDAVSLQKALIEHQLGVPPTSXSFVSGRGENTLPLEVRGAMVIRV 180

Db      180 NSLTRGHSANRLVLEALTNLNHRITPIVPLRGISASAGDLSPLSYIAGAITGHPDV 239
181 NSLTRGHSANRLVLEALTNLNHRITPIVPLRGISASAGDLSPLSYIAAITGHPDSKV 240

QY      240 HYLHGETKTFAPARAISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHDSHMLSL 239
241 HXHEGKEXIKKARAIALFGLPEVVLGPKGGLGVNGTAVASAMATLALHDAHMLSL 300

QY      300 QALTALTYEAMVGQGSFAPFIHDVCRPHQGVAVANIRITLSSGSAFVHEHEEVK 359
301 QALTALTYEAMVGAGSHHPFLHDVTRHPQGLEAVANIRITLSSGSAFVHEHEEVK 360

Db      360 DEGIIRQDRYPLRTSPQPLGVLVEDMMAVSTLSL-NTTTDPLLDVENKQTAHGNF 418
361 DEGIIRQDRYPLRTSPQPLGVLVEDMMAVSTLSL-NTTTDPLLDVENKQTAHGNF 420

QY      419 QASAVISMEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 478
421 QASAVXNTEKTRIALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 480

Db      479 IAAVASELGHLANPVTTFVQPAEMNGQAVNSIALISARTEANDVLSLLASHLYCTIQ 538
481 AAAYTSELGHLANPVTTFVQPAEMNGQAVNSIALISARTEANDVLSLLASHLYCTIQ 540

QY      539 AVDLRAMELDFKKQDPLPTLLOQHJGTGLDVNA---LALVYKXKLNKRLLEQTTTYDL 594
```

Db	541	AVDLRMEEFKQFPYPPXXXXLLXQHFQXXXTXXXXXXXELKXKXKXKXLEQDINSYDL	6000
Qy	595	EPPEMDFAFYAIGTIVTELLSSBSPANVTITAVANAKVASAEKASLITFEVNRFWQTPSS	6544
Db	601	EPMDHAFSAATGTVEEXLSSXXXXXSAAANAKVXAAEKASLITFXVXXFXWAPSS	6600
Qy	655	QAQAHAYLSPTRTVLVSFYREELGYQARRGDVFWCOETIGSNRSRIYEAIKDGRANHV	7144
Db	661	SSPALXYLSPTRTVLVSFYREELGYQARRGDVFLKQEVITIGSNRSRIYEAIKGRINXV	7200
Qy	715	LVYKQLA 720	
Db	721	LVYKQLA 726	

```

RESULT 5
PCT-US02-26934A-30
: Sequence 30, Application PC/TUS0226934A
: GENERAL INFORMATION:
: APPLICANT: PCBU Services, Inc. et al.
: TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
: TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
: TITLE OF INVENTION: Using Same
: FILE REFERENCE: 29479/50NSCA PCT
: CURRENT APPLICATION NUMBER: PCT/US02/26934A
: PRIOR FILING DATE: 2002-08-23
: PRIOR APPLICATION NUMBER: US 09/939,408
: PRIOR FILING DATE: 2001-08-24
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 30
: LENGTH: 720
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Preferred
: OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (5)..(5)
: OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (12)..(12)
: OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (16)..(16)
: OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (17)..(17)
: OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (19)..(19)
: OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (20)..(20)
: OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (25)..(25)
: OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (27)..(27)
: OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (28)..(28)

```

OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
FEATURE:
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
FEATURE:
NAME/KEY: misc feature
LOCATION: (150)..(150)
OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
FEATURE:
NAME/KEY: misc feature
LOCATION: (153)..(153)
OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
FEATURE:
NAME/KEY: misc feature
LOCATION: (154)..(154)
OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (157)..(157)
OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
FEATURE:
NAME/KEY: misc feature
LOCATION: (159)..(159)
OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
FEATURE:
NAME/KEY: misc feature
LOCATION: (183)..(183)
OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
FEATURE:
NAME/KEY: misc feature
LOCATION: (223)..(223)
OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
FEATURE:
NAME/KEY: misc feature
LOCATION: (225)..(225)
OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
FEATURE:
NAME/KEY: misc feature
LOCATION: (237)..(237)
OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala, Val.
FEATURE:
NAME/KEY: misc feature
LOCATION: (239)..(239)
OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
FEATURE:
NAME/KEY: misc feature
LOCATION: (241)..(241)
OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
FEATURE:
NAME/KEY: misc feature
LOCATION: (242)..(242)
OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
FEATURE:
NAME/KEY: misc feature
LOCATION: (246)..(246)
OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
FEATURE:
NAME/KEY: misc feature
LOCATION: (251)..(251)
OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
FEATURE:
NAME/KEY: misc feature
LOCATION: (259)..(259)
OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
FEATURE:
NAME/KEY: misc feature
LOCATION: (305)..(305)
OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
NAME/KEY: misc feature

Query Match 75.6%; Score 2730; DB 1; Length 720;
Best local similarity 76.5%; Pred. No. 3,7e-258;
Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;
QY 1 MAPSDSLATTLANGFTNGSHAPKGAAPTSA LRTPGLDGHAAHQSLIYQELISD 60
DB 1 MAPSDSLATTLANGFTNGSHAPKGAAPTSA LRTPGLDGHAAHQSLIYQELISD 60
QY 61 PDDVVEISGVSLTRVDVVGAAKGRPRVQNDDEIFARVDKSDVDFPKAQDQNSVQVT 120
DB 61 PDDVVEISGVSLTRVDVVGAAKGRPRVQNDDEIFARVDKSDVDFPKAQDQNSVQVT 120
QY 121 GFGGADTRTEDAVSLQALIEHQ:CGVTPTSXSSFSVGRGLENTLEPEVRGAMVRN 180
DB 121 GFGGADTRTEDAVSLQALIEHQ:CGVTPTSXSSFSVGRGLENTLEPEVRGAMVRN 180
QY 181 SLTRGSAVRLVVEALTNELNRITPIVPRGSIASGSLSPSTYAGAITGHPDKV 240
DB 181 SLTRGSAVRLVVEALTNELNRITPIVPRGSIASGSLSPSTYAGAITGHPDKV 240
QY 241 VLHCEKIMFARERASLPLGLEAVVLGPKESLGLVNGTAVSASMATSLDSHMLSLSQ 300
DB 241 VLHCEKIMFARERASLPLGLEAVVLGPKESLGLVNGTAVSASMATSLDSHMLSLSQ 300
QY 301 ALTALTVEAMVGOQSPAPFIHDVCRPHPGQVEYARNRITLLSSSPAVHEBEVKYKD 360
DB 301 ALTALTVEAMVGOQSPAPFIHDVCRPHPGQVEYARNRITLLSSSPAVHEBEVKYKD 360
QY 361 EGLIKQDYPIRTSPQFGLPVEDMMEAVSTLSENNTTNDPLDVENQOTAHGNGFOA 420
DB 361 EGLIKQDYPIRTSPQFGLPVEDMMEAVSTLSENNTTNDPLDVENQOTAHGNGFOA 420
QY 421 SAVSISMETRLALALIGKLFNFTQCTELNNAAMNRLPSCIAEDPSLNYHGKGLDIIHA 480
DB 421 SAVSISMETRLALALIGKLFNFTQCTELNNAAMNRLPSCIAEDPSLNYHGKGLDIIHA 480
QY 481 AYASLGHLANPVTTTFOPAEMGNOAVNSLALISARPTAANDVLSLLASHYCTIQAV 540
DB 481 AYASLGHLANPVTTTFOPAEMGNOAVNSLALISARPTAANDVLSLLASHYCTIQAV 540
QY 541 DLRAMELDFKKQFPLFTLLQHLGTGLDVNALALVKKALNRLBQTTYDLEPRMHD 600
DB 541 DLRAMELDFKKQFPLFTLLQHLGTGLDVNALALVKKALNRLBQTTYDLEPRMHD 600
QY 601 AFSVATGVVELLSSSPEANVTITAVNAKVASAEKALISLREYRNFQWTPSSQADPAHA 660
DB 601 AFSVATGVVELLSSSPEANVTITAVNAKVASAEKALISLREYRNFQWTPSSQADPAHA 660
QY 661 YLSPRTVLVSFVREBELGVKARRGDVFLGQEVITGNVSRITIEALIKGXINHVLVWMLA 720
DB 661 YLSPRTVLVSFVREBELGVKARRGDVFLGQEVITGNVSRITIEALIKGXINHVLVWMLA 720
RESULT 6
US-09-939-408a-30
Sequence 30, Application US/09939408A
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939, 408A
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624, 693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 30
LENGTH: 720

```
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Preferred
OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
NAME/KEY: misc feature
LOCATION: (5)..(15)
OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (12)..(12)
OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (16)..(16)
OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (17)..(17)
OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (19)..(19)
OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
NAME/KEY: misc feature
LOCATION: (20)..(20)
OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
NAME/KEY: misc feature
LOCATION: (25)..(25)
OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (27)..(27)
OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (28)..(28)
OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
NAME/KEY: misc feature
LOCATION: (34)..(34)
OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
NAME/KEY: misc feature
LOCATION: (36)..(36)
OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (39)..(39)
OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (40)..(40)
OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
NAME/KEY: misc feature
LOCATION: (48)..(48)
OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala, Val
NAME/KEY: misc feature
LOCATION: (54)..(54)
OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
NAME/KEY: misc feature
LOCATION: (56)..(56)
OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
NAME/KEY: misc feature
LOCATION: (65)..(65)
OTHER INFORMATION: The 'Xaa' at location 65 stands for Gly, Asp, or Val.
NAME/KEY: misc feature
LOCATION: (66)..(66)
OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (76)..(76)
OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
NAME/KEY: misc feature
LOCATION: (87)..(87)
OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (93)..(93)
OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
NAME/KEY: misc feature
LOCATION: (102)..(102)
OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
NAME/KEY: misc feature
LOCATION: (103)..(103)
OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (109)..(109)
OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (112)..(112)
OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
NAME/KEY: misc feature
LOCATION: (114)..(114)
OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (117)..(117)
OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
NAME/KEY: misc feature
LOCATION: (150)..(150)
OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
NAME/KEY: misc feature
LOCATION: (153)..(153)
OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
NAME/KEY: misc feature
LOCATION: (154)..(154)
OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
NAME/KEY: misc feature
LOCATION: (157)..(157)
OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
NAME/KEY: misc feature
LOCATION: (159)..(159)
OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
NAME/KEY: misc feature
LOCATION: (183)..(183)
OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
NAME/KEY: misc feature
LOCATION: (223)..(223)
OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
NAME/KEY: misc feature
LOCATION: (225)..(225)
OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
NAME/KEY: misc feature
LOCATION: (237)..(237)
OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala.
NAME/KEY: misc feature
LOCATION: (239)..(239)
OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
NAME/KEY: misc feature
LOCATION: (241)..(241)
OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
NAME/KEY: misc feature
LOCATION: (242)..(242)
OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (259)..(259)
OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (305)..(305)
OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
NAME/KEY: misc feature
LOCATION: (319)..(319)
OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
NAME/KEY: misc feature
LOCATION: (346)..(346)
OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
NAME/KEY: stop codon, Trp, or Ser.
```



```

/ LOCATION: (411)..(411)
/ OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
/ NAME/KEY: misc feature
/ LOCATION: (421)..(421)
/ OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
/ NAME/KEY: misc feature
/ LOCATION: (457)..(457)
/ OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
/ NAME/KEY: misc feature
/ LOCATION: (458)..(458)
/ OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
/ NAME/KEY: misc feature
/ LOCATION: (466)..(466)
/ OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
/ NAME/KEY: misc feature
/ LOCATION: (487)..(487)
/ OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
/ NAME/KEY: misc feature
/ LOCATION: (493)..(493)
/ OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
/ NAME/KEY: misc feature
/ LOCATION: (518)..(518)
/ OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
/ NAME/KEY: misc feature
/ LOCATION: (536)..(536)
/ OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
/ NAME/KEY: misc feature
/ LOCATION: (556)..(556)
/ OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
/ NAME/KEY: misc feature
/ LOCATION: (557)..(557)
/ OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
/ NAME/KEY: misc feature
/ LOCATION: (558)..(558)
/ OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.
/ NAME/KEY: misc feature

```

Query Match 75.6%; Score 2730; DB 24; Length 720;
 Best Local Similarity 76.5%; Pred. No. 3.7e-256;
 Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

```

1 MAPSLDLSLATTANGFTNGSHAAPTKSAAGPTSAALRRPGLDGHAAHQSOUEIYOELISD 60
1 MAPSXDSISHTANGXNXXHAPKXXXTGATSLXLRPXLAPRATQTDIDIXLILAD 60
61 PDDVVELSGSYLTVRDVVGARKGRVRVQNDDEIRAVDKSVDFLKAQOLNSVGYTT 120
61 PDDDXLELDGYTLTXDVVGARKGRVRVQTXDEIRAKIDXVEFLFXQLXNVVYVTT 120
121 GFGSADRTTEPAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMVIRVN 180
121 GFGSADRTTEPAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMVIRVN 180
181 SLTRGSAVRLVLEALTNFNLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
181 SLTRGSAVRLVLEALTNFNLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
241 VHEGTEKIMFARERALSIFGLNAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 VHEGTEKIMFARERALSIFGLNAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 XHHEGEXEKIMXAEALALGLPEVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
301 ALTALTVEAMVQGGSPAFPIHDVCRPHPGQVEVARNRITLLSGSFVHEEVEVKYKD 360
301 ALTALTVEAMVQGGSPAFPIHDVCRPHPGQVEVARNRITLLSGSFVHEEVEVKYKD 360
361 EGLLRDRPLRTSPQFLPVEDMMAHAYSTLSLNNNTTTPNPLIDVENKQTAHGNFQA 420
361 EGLLRDRPLRTSPQFLPVEDMMAHAYSTLSLNNNTTTPNPLIDVENKQTAHGNFQA 420
421 SAVSISMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 480
421 SAVSISMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 480
421 XAVANMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 480

```

```

QY 481 AYASELGHIANFVTTFVQPAEMGNQAVNSLALISARPTAENDVLSLLASHLYCTQAV 540
DB 481 AYASELGHIANFVTTFVQPAEMGNQAVNSLALISARPTAENDVLSLLASHLYCTQAV 540
QY 541 DIFAMELDRKQPDPLPPLLQOHGTGIDVNAALAEVKKALNKRLEOTTTYYDIEPRVHD 600
DB 541 DIFAMELDRKQPDPLPPLLQOHGTGIDVNAALAEVKKALNKRLEOTTTYYDIEPRVHD 600
QY 601 AFSVATGTVEILLSSSPSANVTU7VNAAMKVASAEKASITFREYANRFWQTPSSQAPAH 660
DB 601 AFSVATGTVEILLSSSPSANVTU7VNAAMKVASAEKASITFREYANRFWQTPSSQAPAH 660
QY 661 YLSPRTVLYSFVREELGVQARQGVFVQOETIGSNVSSIYEAIKGRINHVLYKXLA 720
DB 661 YLSPRTVLYSFVREELGVQARQGVFVQOETIGSNVSSIYEAIKGRINHVLYKXLA 720

```

RESULT 7
 PCT-US02-26934A-19
 / Sequence 19, Application PC/TUS0226934A
 / GENERAL INFORMATION:
 / APPLICANT: PCB Services, Inc. et al.
 / TITLE OF INVENTION: Phenylalanine Ammonia lyase Polypeptide and
 / TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 / TITLE OF INVENTION: Using Same
 / FILE REFERENCE: 29479/50NSCA PCT
 / CURRENT APPLICATION NUMBER: PCT/US02/26934A
 / PRIOR FILING DATE: 2002-08-23
 / PRIOR APPLICATION NUMBER: US 09/939,408
 / NUMBER OF SEQ ID NOS: 30
 / SOFTWARE: Patentln Ver. 2.0
 / SEQ ID NO 19
 / LENGTH: 716
 / TYPE: PRT
 / ORGANISM: Rhodocorula toruloides
 PCT-US02-26934A-19

Query Match 72.4%; Score 2614; DB 1; Length 716;
 Best Local Similarity 72.0%; Pred. No. 9.5e-247;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSLATTANGFTNGSHAAPTKSAAGPTSAALRRPGLDGHAAHQSOUEIYOELISD 60
1 MAPSLDLSHSTANG-----VASAKQAVNGASTNLAAGSHLPTTQVTVQDIYERKMLAA 54
61 PDDVVELSGSYLTVRDVVGARKGRVRVQNDDEIRAVDKSVDFLKAQOLNSVGYTT 120
61 PDDDXLELDGYTLTXDVVGARKGRVRVQTXDEIRAKIDXVEFLFXQLXNVVYVTT 120
121 GFGSADRTTEPAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMVIRVN 180
121 GFGSADRTTEPAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMVIRVN 180
181 SLTRGSAVRLVLEALTNFNLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
181 SLTRGSAVRLVLEALTNFNLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
241 VHEGTEKIMFARERALSIFGLNAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 VHEGTEKIMFARERALSIFGLNAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 XHHEGEXEKIMXAEALALGLPEVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
301 ALTALTVEAMVQGGSPAFPIHDVCRPHPGQVEVARNRITLLSGSFVHEEVEVKYKD 360
301 ALTALTVEAMVQGGSPAFPIHDVCRPHPGQVEVARNRITLLSGSFVHEEVEVKYKD 360
361 EGLLRDRPLRTSPQFLPVEDMMAHAYSTLSLNNNTTTPNPLIDVENKQTAHGNFQA 420
361 EGLLRDRPLRTSPQFLPVEDMMAHAYSTLSLNNNTTTPNPLIDVENKQTAHGNFQA 420
421 SAVSISMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 479
421 SAVSISMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 479
421 XAVANMEKTRIALALIGKLNFTQCTELNLNAMRGSLAAEDPSLNHYGGLDIIH 479

```

QY 480 AAVASELGHLANPVTTFFVQPAEMNGQAVNSLALISARRTEANDVLSLLASHLYCTLOA 539
DB 475 AAVTSELGHLANPVTTTHQPAEMANQAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDFKKQFDPPLPTLLQOHLG---TGLDV-NALALEVKKALKKLEQTTYYLE 595
DB 535 IDLRALIEFEFKQFQGPALVSLIDQHPGAMTGSNLRDELVEKVNKTAKLEQTNNSYDLV 594
QY 596 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVASAEKALSTREVRNRFMQPSSQ 655
DB 595 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVASAEKALSTREVRNRFMQPSSQ 651
QY 656 APAAHAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRLEYAIKQGRINHL 715
DB 652 SPALSYLSPTRTQILVAFVREELGVKARRGDVFLGKQEVITIGSNVSKLEYAIKGRINHL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 8
PCT-US03-05708-2
; Sequence 2, Application PC/TUS0305708
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours and Company, Inc.
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: C11794 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/05708
; PRIOR FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
PCT-US03-05708-2

Query Match 72.4%; Score 2614; DB 1; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 1 MAPSLDLSLATTLANGFTNGSHAPTKSAGPTSA LRTPGIDGHAHQSOLEIYQELISD 60
DB 1 MAPSLDLSLHSPFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQDIVEKMLAA 54
QY 61 PTDDVVELSGVSLTVRDVVGAAKGRVRVQNDDEIRAVDKSVDFPKAQLQNSVYGVTT 120
DB 55 PTDSLTLEDGYSLNIGDVVSAARKGRPVAVKDSDEIRSKIDKSVFELRSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTXSSGSFVGRGLENTLPLEVRGAMVIRVN 180
DB 115 GFGGSADRTEDAVSLQKALIEHQLCGVLPSPFSFRLGRGLENTLPLEVRGAMVIRVN 174
QY 181 SLTRGHSARLVVLEALTNFLNHRITPIVPLRGISASGDSLPLSYIAGAITGHPDVYH 240
DB 175 SLTRGHSARLVVLEALTNFLNHRITPIVPLRGISASGDSLPLSYIAGAITGHPDVYH 234
QY 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKEGIGLVNGTAVASASMAATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLBEPVVLGPKEGIGLVNGTAVASASMAATLSLHSHMLSLISQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGOVEVYARNRITLSSGSPFVHEEVEVKYKD 360
DB 295 SLTANTVEAMVGHASFPFIHDVTRPHPTQIEVAGNIRKLLBSGRFAVHEEVEVKYKD 354
QY 361 EGI LRQDRYPLRTSPQFGLPVEDMGMAYSTLSLE-NNTTTDNPLLDVENKQTAHGNFQ 419
DB 355 EGI LRQDRYPLRTSPQFGLPVEDMGMAYSTLSLE-NNTTTDNPLLDVENKQTAHGNFQ 414

QY 420 AAVSISMEKTRILALALGKLNFTQCTELTANANRGLPSCLAEDPSLNYHGKGLDIH 479
DB 415 AAVANVMEKTRILGALQIGKLNFTQCTELTANANRGLPSCLAEDPSLNYHGKGLDIAA 474
QY 480 AAVASELGHLANPVTTFFVQPAEMNGQAVNSLALISARRTEANDVLSLLASHLYCTLOA 539
DB 475 AAVTSELGHLANPVTTTHQPAEMANQAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDFKKQFDPPLPTLLQOHLG---TGLDV-NALALEVKKALKKLEQTTYYLE 595
DB 535 IDLRALIEFEFKQFQGPALVSLIDQHPGAMTGSNLRDELVEKVNKTAKLEQTNNSYDLV 594
QY 596 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVASAEKALSTREVRNRFMQPSSQ 655
DB 595 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVASAEKALSTREVRNRFMQPSSQ 651
QY 656 APAAHAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRLEYAIKQGRINHL 715
DB 652 SPALSYLSPTRTQILVAFVREELGVKARRGDVFLGKQEVITIGSNVSKLEYAIKGRINHL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 9
PCT-US03-17926-2
; Sequence 2, Application PC/TUS0317926
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: C11912 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/17926
; PRIOR FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
PCT-US03-17926-2

Query Match 72.4%; Score 2614; DB 1; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 1 MAPSLDLSLATTLANGFTNGSHAPTKSAGPTSA LRTPGIDGHAHQSOLEIYQELISD 60
DB 1 MAPSLDLSLHSPFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQDIVEKMLAA 54
QY 61 PTDDVVELSGVSLTVRDVVGAAKGRVRVQNDDEIRAVDKSVDFPKAQLQNSVYGVTT 120
DB 55 PTDSLTLEDGYSLNIGDVVSAARKGRPVAVKDSDEIRSKIDKSVFELRSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGVTPTXSSGSFVGRGLENTLPLEVRGAMVIRVN 180
DB 115 GFGGSADRTEDAVSLQKALIEHQLCGVLPSPFSFRLGRGLENTLPLEVRGAMVIRVN 174
QY 181 SLTRGHSARLVVLEALTNFLNHRITPIVPLRGISASGDSLPLSYIAGAITGHPDVYH 240
DB 175 SLTRGHSARLVVLEALTNFLNHRITPIVPLRGISASGDSLPLSYIAGAITGHPDVYH 234
QY 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKEGIGLVNGTAVASASMAATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLBEPVVLGPKEGIGLVNGTAVASASMAATLSLHSHMLSLISQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGOVEVYARNRITLSSGSPFVHEEVEVKYKD 360
DB 295 SLTANTVEAMVGHASFPFIHDVTRPHPTQIEVAGNIRKLLBSGRFAVHEEVEVKYKD 354
QY 361 EGI LRQDRYPLRTSPQFGLPVEDMGMAYSTLSLE-NNTTTDNPLLDVENKQTAHGNFQ 419
DB 355 EGI LRQDRYPLRTSPQFGLPVEDMGMAYSTLSLE-NNTTTDNPLLDVENKQTAHGNFQ 414

```

Db      355  EQLLQDRIPLRTSPQWLGPIVSDLIHAAVLTIEAGOSTTNDPLIDVENKTHGGNQ 414
Qy      420  ASAVSISNEKTRALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 479
Db      415  AAAVANTVEKTRGLAQIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 474
Qy      480  AAYASEIGHLANPVTTFQPAEMNQAVNSLALISARTAEANDVLSLLASHTYCTLOA 539
Db      475  AAYTSELGHLANPVTTFQPAEMNQAVNSLALISARTAESNDVLSLLATLHYCVLOA 534
Qy      540  VDLRAMELDFKQDPPLPTLLQOHLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db      535  IDLRAIEEPFKQGPALIVSLIDHFGSAMTGSNLRDELVEKNKTLAKRLQETNSYDLV 594
Qy      556  PRMHDAFSAATGTVELLSPPSANVTTLTANNAKVASAEKALSTREVRNRPWQTPSSQ 655
Db      595  PRMHDAFSAATGTVEVLST---SLSLAANNAKVAASASALSTRQVRETWSAAS 651
Qy      656  APAHAYLSPTRTVLYSPFREELGVQARRGVFVGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db      652  SPALSYLSPTRTQILYAFVREELGVKARRGVFLGKQEVITIGSNVSKITYEAIKGRINNV 711
Qy      716  VKMLA 720
Db      712  LKMLA 716

RESULT 10
US-09-939-408a-19
/ Sequence 19, Application US/09939408A
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ APPLICANT: Koester, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ FILE REFERENCE: 29479/500NSCA
/ CURRENT APPLICATION NUMBER: US/09/939, 408A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 09/624, 693
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 19
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Rhodotorula toruloides
JS-09-939-408a-19

Query Match      72.4%, Score 2614, DB 24, Length 716,
Best Local Similarity 72.0%, Pred. No. 9.5e-247;
Matches 522, Conservative 86, Mismatches 103, Indels 14, Gaps 5;

Db      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAIRRTPGLDGHAHQSLQLEIVGELLSD 60
Qy      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAIRRTPGLDGHAHQSLQLEIVGELLSD 60
Db      55  PTDSTLELDGSLNLDGVVSAARKGRPVKDSDBIRSKIDKSVFLRSQLSNVSYGVTT 114
Qy      55  PTDSTLELDGSLNLDGVVSAARKGRPVKDSDBIRSKIDKSVFLRSQLSNVSYGVTT 114
Db      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVRGAMTIRVN 180
Qy      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVRGAMTIRVN 180
Db      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSFRSLRGLENSLPLEVVRGAMTIRVN 174
Qy      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSFRSLRGLENSLPLEVVRGAMTIRVN 174
Db      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSYAAGITGHPVKVN 240
Qy      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSYAAGITGHPVKVN 240
Db      175  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGTISAGDSLPLSYAAGITGHPVKVN 234
Qy      175  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGTISAGDSLPLSYAAGITGHPVKVN 234
Db      241  VLHGETEKIMPARBAISLFGLEAVVLGPKEGGLGVNQTAVASAMATLSLDSHMLLSLQ 300
Qy      241  VLHGETEKIMPARBAISLFGLEAVVLGPKEGGLGVNQTAVASAMATLSLDSHMLLSLQ 300

```

```

Db      225  VHEGKEKILTAAREAMALFNIEBPVVLGPKEGGLGVNQTAVASAMATLSLDSHMLLSLQ 294
Qy      301  ALTAITYEAMVQOQSPAPFIHDVCRPHQOVEARINRITLSSGFAYHEEEVKKVD 360
Db      225  SLTAMTYEAMVGHAGSPHFLHDVTRPHQIIEVAGNIRKLGLSGSPAYHEEEVKKVD 354
Qy      361  EGLIRQDRYPLRTSPQWLGPIVSDLIHAAVLTIEAGOSTTNDPLIDVENKTHGGNQ 419
Db      355  EGLIRQDRYPLRTSPQWLGPIVSDLIHAAVLTIEAGOSTTNDPLIDVENKTHGGNQ 414
Qy      420  ASAVSISNEKTRALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 479
Db      415  AAAVANTVEKTRGLAQIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 474
Qy      480  AAYASEIGHLANPVTTFQPAEMNQAVNSLALISARTAEANDVLSLLASHTYCTLOA 539
Db      475  AAYTSELGHLANPVTTFQPAEMNQAVNSLALISARTAESNDVLSLLATLHYCVLOA 534
Qy      540  VDLRAMELDFKQDPPLPTLLQOHLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db      535  IDLRAIEEPFKQGPALIVSLIDHFGSAMTGSNLRDELVEKNKTLAKRLQETNSYDLV 594
Qy      556  PRMHDAFSAATGTVELLSPPSANVTTLTANNAKVASAEKALSTREVRNRPWQTPSSQ 655
Db      595  PRMHDAFSAATGTVEVLST---SLSLAANNAKVAASASALSTRQVRETWSAAS 651
Qy      656  APAHAYLSPTRTVLYSPFREELGVQARRGVFVGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db      652  SPALSYLSPTRTQILYAFVREELGVKARRGVFLGKQEVITIGSNVSKITYEAIKGRINNV 711
Qy      716  VKMLA 720
Db      712  LKMLA 716

RESULT 11
US-10-138-970A-4
/ Sequence 4, Application US/10138970A
/ GENERAL INFORMATION:
/ APPLICANT: Sarlasani, Sima
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei-Wei
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ TITLE OF INVENTION: hydroxycinnamic Acid
/ FILE REFERENCE: CL1777
/ CURRENT APPLICATION NUMBER: US/10/138, 970A
/ PRIOR FILING DATE: 2002-07-23
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 4
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: C. violaceum
US-10-138-970A-4

Query Match      72.4%, Score 2614, DB 27, Length 716,
Best Local Similarity 72.0%, Pred. No. 9.5e-247;
Matches 522, Conservative 86, Mismatches 103, Indels 14, Gaps 5;

Db      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAIRRTPGLDGHAHQSLQLEIVGELLSD 60
Qy      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAIRRTPGLDGHAHQSLQLEIVGELLSD 60
Db      55  PTDSTLELDGSLNLDGVVSAARKGRPVKDSDBIRSKIDKSVFLRSQLSNVSYGVTT 114
Qy      55  PTDSTLELDGSLNLDGVVSAARKGRPVKDSDBIRSKIDKSVFLRSQLSNVSYGVTT 114
Db      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVRGAMTIRVN 180
Qy      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVRGAMTIRVN 180
Db      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSFRSLRGLENSLPLEVVRGAMTIRVN 174
Qy      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSFRSLRGLENSLPLEVVRGAMTIRVN 174
Db      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSYAAGITGHPVKVN 240
Qy      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSYAAGITGHPVKVN 240

```

```

DB 175 SLTRGSAVRLVLEALTNFLNHTGTPVLPRLGRTISASGDSLPSIYIAAIIISGHPDSKVH 224
QY 241 VVHEGTEKIMFAREAISLFGLEAVVLGPKREGGLVNGTAVSASMAATLSLHDSHMLLSLQ 300
DB 235 VVHEGTEKILYAREAMALFNLEPVVLGPKREGGLVNGTAVSASMAATLSLHDSHMLLSLQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGQVEVAANIRTLTSSGSAFAHEEVEVKXD 360
DB 295 SLTAMTEAVMVGAGSHFPHLDVTRPHPTQIEVAANIRKLEGSFAFVHEEVEVKXD 354
QY 361 EGLIRDRYPLRTSPQWLGPLVSDLIHAHVLITIEAGOSTTDNPLIDVENKTSHHGQNFQ 414
DB 355 EGLIRDRYPLRTSPQWLGPLVSDLIHAHVLITIEAGOSTTDNPLIDVENKTSHHGQNFQ 414
QY 420 ASAVSISMEKTRALALIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
DB 415 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 480 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
DB 475 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 540 VDLRAVELDKKQDPPLPTLLOHIG--TGLDV-NALALEYKALNRELEQTTTYDLE 595
DB 535 IDLRALIEFEFKQGPALIVSLIDQHFGSAMTGNLDELVEKVKTKLAKLEQTNSTYDLY 594
QY 596 PRMHDAFSYATGTVEVLELSSSPSANTTLTAVNAKVAASAKAISLITREVRNFMQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLELSSSPSANTTLTAVNAKVAASAKAISLITREVRNFMQTPSSQ 655
QY 656 APAAHAYLSPTRTVLISFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIHNL 715
DB 652 SPALSYLSPRTQILVAFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIHNL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 12
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

Query Match 72.4%; Score 2614; DB 27; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

```

QY 1 MAPSDLSATTLANGFTNGSHAAPTSAARPTGDLGHAHQSGLIYQELLSD 60
DB 1 MAPSDLSATTLANGFTNGSHAAPTSAARPTGDLGHAHQSGLIYQELLSD 54
QY 61 PDDVYVELISGVSLTRDVVGAARKGRVRYONDEIDARVADKSDPLKALQKSVYGVTT 120
DB 55 PDDVYVELISGVSLTRDVVGAARKGRVRYONDEIDARVADKSDPLKALQKSVYGVTT 114

```

```

QY 121 GFGSADTRTEDAYASLOKALIEHOLGVPTXSFSVGRGLENTLPLEVVGAMVIRVN 180
DB 115 GFGSADTRTEDAYASLOKALIEHOLGVPTXSFSVGRGLENTLPLEVVGAMVIRVN 174
QY 181 SLTRGSAVRLVLEALTNFLNHTGTPVLPRLGRTISASGDSLPSIYIAAIIISGHPDSKVH 240
DB 175 SLTRGSAVRLVLEALTNFLNHTGTPVLPRLGRTISASGDSLPSIYIAAIIISGHPDSKVH 234
QY 241 VVHEGTEKIMFAREAISLFGLEAVVLGPKREGGLVNGTAVSASMAATLSLHDSHMLLSLQ 300
DB 235 VVHEGTEKILYAREAMALFNLEPVVLGPKREGGLVNGTAVSASMAATLSLHDSHMLLSLQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGQVEVAANIRTLTSSGSAFAHEEVEVKXD 360
DB 295 SLTAMTEAVMVGAGSHFPHLDVTRPHPTQIEVAANIRKLEGSFAFVHEEVEVKXD 354
QY 361 EGLIRDRYPLRTSPQWLGPLVSDLIHAHVLITIEAGOSTTDNPLIDVENKTSHHGQNFQ 414
DB 355 EGLIRDRYPLRTSPQWLGPLVSDLIHAHVLITIEAGOSTTDNPLIDVENKTSHHGQNFQ 414
QY 420 ASAVSISMEKTRALALIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
DB 415 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 480 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
DB 475 AAVALNTEKTRGLAIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 540 VDLRAVELDKKQDPPLPTLLOHIG--TGLDV-NALALEYKALNRELEQTTTYDLE 595
DB 535 IDLRALIEFEFKQGPALIVSLIDQHFGSAMTGNLDELVEKVKTKLAKLEQTNSTYDLY 594
QY 596 PRMHDAFSYATGTVEVLELSSSPSANTTLTAVNAKVAASAKAISLITREVRNFMQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLELSSSPSANTTLTAVNAKVAASAKAISLITREVRNFMQTPSSQ 655
QY 656 APAAHAYLSPTRTVLISFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIHNL 715
DB 652 SPALSYLSPRTQILVAFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIHNL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 13
US-10-188-523C-8
; Sequence 8, Application US/10188523C
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009-CIP
; CURRENT APPLICATION NUMBER: US/10/188,523C
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-188-523C-8

Query Match 72.4%; Score 2614; DB 27; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

```

QY 1 MAPSDLSATTLANGFTNGSHAAPTSAARPTGDLGHAHQSGLIYQELLSD 60
DB 1 MAPSDLSATTLANGFTNGSHAAPTSAARPTGDLGHAHQSGLIYQELLSD 54

```

6 PTDDVVLSSYSTVDDVVGGAARKGRVAVONDDDEIRARVVKSDPLKAOQNSVGVTT 120
 55 PTDSLELDGYSINLGDVVAARKGRPRVAVKSDSEIRKIRKSVFELLSQNSVGVTT 114
 121 GFGGSADPTEDAVSIQKALIEHQLCGTTPTSXSSFSVGRGLIENTLPLEVRGAMVIRN 160
 115 GFGGSADPTEDDAISIQKALLEHQLCGVLPSSFSDFRIGRGLENSPLLEVRGAMTIRVN 174
 181 SLRGRSAVAVLYLALTNFLNHRITPIYPLRGSIASAGDLSPLSYAGAITGHPDYKH 240
 175 SLRGRSAVAVLYLALTNFLNHRITPIYPLRGITISAGDLSPLSYAAISGHPDSKH 234
 241 VLHEGTEKIMFAEEALSLEGLAVYLGPXEGSLGVNGTAVASAMATLSLHDSHYLSSLG 300
 235 VHEGGEKELIYAEENALFNLEPVLGPXEGSLGVNGTAVASAMATLADHAHYLSLGG 294
 301 ALVALVEAMVQGGSGFAPFHIDVCRPHPGQVEVARNRTILSSGSFVHEEVRVKXD 360
 295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHTQIEVAGNIRKLLBGSRRVAVHEEVRVKXD 354
 361 EGIIRDORYELRTSPQFLPGLVEDMMAHASTLSE--NNTTNDNPLDVENKOTAHGNFQ 419
 355 EGIIRDORYELRTSPQFLPGLVEDMMAHASTLSE--NNTTNDNPLDVENKOTAHGNFQ 414
 420 ASAVSISMEKTRIALALIKLNFQCTELLNAANRGLPSCLAEDPSLNYHGKLDIHI 479
 415 AAAVANMEXTRIGLAIQIKLNFQCTELLNAANRGLPSCLAEDPSLNYHGKLDIHI 474
 480 AAASSELGHANVTTFQVQPAAMGQAVNSLALISARTAEANVLSLLASHLYCTIQ 539
 475 AATSELGHANVTTHVQPAAMANOAVNSLALISARTTESNVLSLLATHLYCVLQ 534
 540 VDIRAMELDFKKQFDELLPTLLIQHLG--TGLDV-NALALEVKKALNKLEQTTIYDE 595
 535 IDIRAEFEFFKQFGPAIVSLIDQFGSAMGNSNRLDELVEKVKTKLAEQNSYDLV 594
 596 PRHADAFSYATGTVVEILLSSPSANVTILTAVAMVVASAEALILTVEYNNRFPWQPSQ 655
 595 PRHADAFSYATGTVVEILLSSPSANVTILTAVAMVVASAEALILTVEYNNRFPWQPSQ 651
 656 APAHAALSPRTVLSFVVEELGVQARRGDVFPVQOCTTISNRSRIYEAIXDGRINHYL 715
 652 SPALSTLSPTQLVAFVVEELGVQARRGDVFPVQOCTTISNRSRIYEAIXDGRINHYL 711
 716 VKKLA 720
 712 LKKLA 716
 RESULT 14
 US-10-374-366-2
 / Sequence 2, Application US/10374366
 / GENERAL INFORMATION:
 / APPLICANT: Tang, Xiao-Song
 / APPLICANT: Milano, Joseph
 / TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
 / FILE REFERENCE: C11794 US NA
 / CURRENT APPLICATION NUMBER: US/10/374,366
 / CURRENT FILING DATE: 2003-02-26
 / PRIOR APPLICATION NUMBER: 60/360,279
 / PRIORITY FILING DATE: 2002-02-26
 / NUMBER OF SEQ ID NOS: 203
 / SOFTWARE: PatentIn version 3.2
 / SEQ ID NO 2
 / LENGTH: 716
 / TYPE: PRT
 / ORGANISM: Rhodospiridium glutinis
 / IS-10-374-366-2

Query Match	72.4%;	Score 2614;	DB 29;	Length 716;
Best Local Similarity	72.0%;	Pred. No. 9.5e-247;		
Matches 522;	Conservative	86;	Mismatches 103;	Indels 14;
				Gaps 5;

```

QY 1 MABSLDLSATTLANGTNSHAPPKYSAAGPFSALRRPPGLDGHAAHOSQLEIYOELLSD 60
Db 1 MABSLDLSISSFANG-----VASAQOANGASTNLAWAGSHLPYTOYDVIYERKMLAA 54
QY 61 PTDDVVELSGSYLTVDVYGAARKGRVYVONDEIRARVDKSVDELFAQONSVYGVTT 120
Db 55 PLSSTLELDGYSINTLDGVDYSAARKGRPARVKOSDIERIKDYSVEFLRSQLSMVSYGTT 112
QY 121 GFQGSADTRTEDAVSIQKALIEHOLCGVTPTISXSFSVGRGIENTLPLEYVRGAMVTRVN 180
Db 115 GFQGSADTRTEDASISQKALLEHOLCGVPSFSFDSFRGRGLENLPLEYVRGAMTRIVN 174
QY 181 SLRGSHAVLYVLEALNTPLNHRITPEIYPLRGSTSAAGDSLPSLSYAGATGTHPDVKVN 240
Db 175 SLRGSHAVLYVLEALNTPLNHRITPEIYPLRGSTSAAGDSLPSLSYAAASGHDPKVN 233
QY 241 VLEHGETKIMFARFASILFLEBAVYLGPKEGIGLVNGTAVASAMATLSLHDSHMLSLSQ 300
Db 235 VAEHGEKEXILYAEAEAMALFNLEPVLGPREGIGLVNGTAVASAMATLALHDAHMLSLSQ 299
QY 301 ALPLALVEAMVGOQSSPAFPIHDVCPHPGQVEVARNIRTLISGSSPAVEHEEVEKYKD 360
Db 295 SLTAMTVEAMVHAGSFHPFLHDVTRPHTQIEVAGNIRKLLGSRFAYHHEEVEKYKD 354
QY 361 EGLTRDQRYVLRTPQFLPYEDMKHASTLSLE-NNTTTDNLPLVYENQZHAAGNFQ 412
Db 355 EGLTRDQRYVLRTPSPWMLPGLVSDLIHAALVLTIEAGOSTTNDPLIDVENTSHHGQNFQ 412
QY 420 ASAVSISMEKTRIALALIGKLNFTOCTEELNANMRGLPSCLAEDPSELNHYGKGLDHI 472
Db 415 AAAYANIMETKTRIGLAQIGKLNFTQUTEMLANGMKGLPSCLAEDPSELNHYGKGLDIA 472
QY 480 AAAYASBELGIANEVTTFVQPAEMGQAVNSLALISARRTAENDVLSILLASHLYCTLOA 533
Db 475 AAYTSELGHLANVTTHTVQPAEMANQAVNSLALISARRTESNDVLSILLATHLYCYLOA 533
QY 540 VDIRAMELDPKQFQDLEPTLLQOHLG---TGLDV-NALALEYKALKLRLEQTTVDE 599
Db 535 IDIRALEFEKKQFGAIVSLIDQHFGSAMGTSNIRDELVERKNTLAKRLEQNSVDLV 594
QY 596 PRMHDAFSYATGTVGVELSSPSAVUTLTAVANMVAEAFASILTEVRARPMQTPSSQ 655
Db 595 PRMHDAFSYAGTVGVELSS---SLSLAANVAMKVAALAEASILTRVRETPFMSAISTS 655
QY 656 APAAHAYISPTRLVLSFYVEELGVQARGDVFVVGQETIGSNVSRLEYAIDKGRINHVL 713
Db 652 SPALSYLSPTQLLYFAVEELGVARGDVFELGQEWETIGSNVSKLEYAISKGRINNVL 711
QY 716 VKXLA 720
Db 712 LKXLA 716

```

```

RESULT 15
US-10-439-478-2
; Sequence 2, Application US/10439478
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arle
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariastani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Varelil, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: C11912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2

```

LENGTH: 716
TYPE: PRT
ORGANISM: Rhodotorula glutinis
US-10-439-478-2

Query Match 72.4%; Score 2614; DB 30; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
QY 1 MAPPSIDSLATTANGFTNGSHAPTKSAAGPTSAIRRTPGIDGHAHQSLBIYQELSD 60  
DB 1 MAPSIDSLISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTTQTVQDIVERKMLAA 54  
QY 61 PDDVVELSGVSLTVRDVYGAARKGRVAVONDEIRARVYKSVDFLKAQLQNSVYGVTT 120  
DB 55 PIDSSTLELDGYSLNIGDVYSAARKGRPRVYKSDSEIRSKIDKSVFRLRSQLSMSYGVTT 114  
QY 121 GFGGSADTRTEDAVSLQALIEHQICGYTPTSXSSFSVGRGLENTLEEVYQAMVIRVN 180  
DB 115 GFGGSADTRTEDAISLQKALLEHQICGYLPSSFSFRLGRGLENTLEEVYRGAMTIRVN 174  
QY 181 SLTRGHSARLVVYVLEALNPNLNRHTPIVPIRGSIASGDLSPSTYAGAITGHPDYKH 240  
DB 175 SLTRGHSARLVVLEALNPNLNRHTPIVPIRGSIASGDLSPSTYAGAITGHPDYKH 234  
QY 241 VHEGTEKIMFARBAISLFGLEAVVLGPKBGLGVNGTAVASAMATLSLHSHMLSLGQ 300  
DB 235 VHEGTEKILYARERAMALFNLEPVYLGPKBGLGVNGTAVASAMATLSLHSHMLSLGQ 294  
QY 301 ALTALTVEAMVQOQSSFAPIHDVCRPHPGOVEVARNIRTLSSSSFAVEHEEEVYKOD 360  
DB 295 SLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHEEEVYKOD 354  
QY 361 EGIQRDRYPLTSPOQFLVEDMWHAYSTLSLE--NNTTNDNPLDVENKOTAHGQNFQ 419  
DB 355 EGIQRDRYPLTSPOQFLVEDMWHAYSTLSLE--NNTTNDNPLDVENKOTAHGQNFQ 414  
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNAAANRQIPSCLAEDPSLNYHGKQDIHI 479  
DB 415 ASAVSISMEKTRIALALIGKLNFTQCTELNAAANRQIPSCLAEDPSLNYHGKQDIHI 474  
QY 480 AAYASBIGHLANPVTTFVQPAEMGNOAVNSIALISARRTAENDVLSLLASHLYCTLOA 539  
DB 475 AAYASBIGHLANPVTTFVQPAEMGNOAVNSIALISARRTAENDVLSLLASHLYCTLOA 534  
QY 540 VDLRAMELDKKQFDPLLPTLLQCHG--TGLDV-NALALEYKALNKRLEQTTTYDLE 595  
DB 535 VDLRAMELDKKQFDPLLPTLLQCHG--TGLDV-NALALEYKALNKRLEQTTTYDLE 594  
QY 596 PRMHDAFSYATGTVELLSSSPSANVTLTANANKVASAEKALSITREVNRFWQTPSSQ 655  
DB 595 PRMHDAFSYATGTVELLSSSPSANVTLTANANKVASAEKALSITREVNRFWQTPSSQ 651  
QY 656 APAHAVALSPRTVLVSFYREELGVQARRGDVFVQOETIGSNVSRITYEAIKDGRIINVL 715  
DB 652 SPALSYLSRTOILYAFREELGVQARRGDVFVQOETIGSNVSRITYEAIKDGRIINVL 711  
QY 716 VMQLA 720  
DB 712 VMQLA 716
```

Search completed: September 9, 2004, 10:19:37
Job time : 298.768 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M protein - protein search, using sw model

Run on: September 9, 2004, 10:04:49 ; Search time 43.8174 Seconds
(without alignments)
2131.280 Million cell updates/sec

Title: US-09-939-408A-13

Perfect score: 3610

Sequence: 1 MAPSLDLSLATTIANGTNGS.....RIEAIKDGRIHVIVKMLA 720

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 704768 seqs, 129704438 residues

Total number of hits satisfying chosen parameters: 704768

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCIT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2614	72.4	716	6	US-10-476-193-4
2	2614	72.4	716	6	US-10-476-198-4
3	2613	72.4	716	6	US-10-476-193-19
4	2613	72.4	716	6	US-10-476-198-19
5	2612	72.4	716	6	US-10-476-193-24
6	2612	72.4	716	6	US-10-476-198-24
7	2607	72.2	716	6	US-10-476-193-22
8	2607	72.2	716	6	US-10-476-198-22
9	2603	72.1	716	6	US-10-476-193-20
10	2603	72.1	716	6	US-10-476-193-21
11	2603	72.1	716	6	US-10-476-198-20
12	2603	72.1	716	6	US-10-476-198-21
13	2599	72.0	716	6	US-10-476-193-18
14	2599	72.0	716	6	US-10-476-198-18
15	2579	71.4	716	6	US-10-476-193-17
16	2579	71.4	716	6	US-10-476-198-17
17	1036	28.7	712	6	US-10-491-823-332
18	1031	28.6	712	6	US-10-491-823-334
19	1010.5	28.0	714	6	US-10-769-350-13
20	1005	27.8	703	6	US-10-425-115-267547
21	1003.5	27.8	671	7	US-60-548-891-28
22	1003.5	27.8	703	6	US-10-425-115-339651
23	999	27.7	716	6	US-10-425-115-218409
24	996.5	27.6	705	6	US-10-767-701-47379
25	993	27.5	714	6	US-10-425-115-218412
26	990	27.4	668	6	US-10-767-701-47377

27	985.5	27.3	713	7	US-60-563-879-184	Sequence 184, App
28	985	27.3	717	6	US-10-425-115-217555	Sequence 217555,
29	983.5	27.2	704	6	US-10-425-115-218411	Sequence 218411,
30	983	27.2	711	6	US-60-563-879-108	Sequence 108, App
31	982	27.2	711	6	US-10-769-350-13	Sequence 13, App1
32	980.5	27.2	713	7	US-60-563-879-109	Sequence 109, App
33	979	27.0	706	6	US-10-425-115-339653	Sequence 339653,
34	977.5	27.0	722	6	US-10-425-115-217553	Sequence 217553,
35	971.5	26.9	672	6	US-10-425-115-339642	Sequence 339642,
36	967	26.8	671	6	US-10-491-823-330	Sequence 330, App
37	966.5	26.8	698	6	US-10-425-115-327033	Sequence 327033,
38	892.5	24.7	619	6	US-10-425-115-217557	Sequence 217557,
39	865.5	24.0	617	6	US-10-425-115-267544	Sequence 267544,
40	854.5	23.7	584	6	US-10-425-115-267542	Sequence 267542,
41	566.5	15.7	499	6	US-10-159-257A-189	Sequence 189, App
42	506.5	14.0	510	7	US-60-579-902-6176	Sequence 6176, App
43	494.5	13.7	277	6	US-10-425-115-267539	Sequence 267539,
44	448.5	12.4	192	6	US-10-491-823-269	Sequence 269, App
45	434.5	12.0	281	6	US-10-425-115-352913	Sequence 352913,

ALIGNMENTS

```
RESULT 1
US-10-476-193-4
; Sequence 4, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sarsisian, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476.193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-476-193-4

Query Match      72.4%; Score 2614; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.2e-194;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1 MAPSLDLSLATTIANGTNGSNAAPTSAARTEGLDGHANQOLEYQELSD 60
      1 MAPSLDISHSFANG-----VASAKQAVNGASTNDLAVAGSHLPTTQVTVQDIVKMLA 54
DB
QY      61 PTDDVVELSGSLTRDVVGAAKGRVRVYQNDDEIRAVDKSVDFLKAQOLQNSVYGVTT 120
      55 PTDTLTLIDGSLNIGVYVAAKGRVRVYKDSDEIRSKIDKSVETLRQOLMSVYGVTT 114
DB
QY      121 GFGSSADTRFDVAISLOKALIEHQLCGVPTFSKXSPSVGRLNTLPLEVVGAMVIRVN 180
      115 GFGSSADTRFDVAISLOKALIEHQLCGVPTFSKXSPSVGRLNTLPLEVVGAMVIRVN 174
DB
QY      161 SLTGHSAVLLVLEALTNLNRITPIVPLRGISASGSLPSTAGATGHPVVKYH 240
      175 SLTGHSAVLLVLEALTNLNRITPIVPLRGISASGSLPSTAGATGHPVVKYH 234
DB
QY      241 VLRGTEKINFARBAISLFLGLAVVLGPKEGLDVGNGTAVSASMATLSLHDSHMLSLSQ 300
      235 VLRGTEKILYAREFAMLFYLEPVVLGPKEGLDVGNGTAVSASMATLSLHDSHMLSLSQ 294
DB
QY      301 ALTKLTLEAMVGGQGFAPFIHYVCRPHRGQVAVANIRTLISGSSFAVHEBEVVKVD 360
      295 SLTAMTVEAMVGAHSGHPFLHDVTRPHLPQLEVAGNIRKTLLEGSRFAVHEBEVVKVD 354
QY      361 EGLRQRVPLRTPQPLGLVEDMNAVSTLSLE-NTTTPDLPLVDENKQTAGHGNFQ 419
```

Db 355 EGIILRODRPLRTSPQWLGPLVSDLIHHAVALTIEAGOSTTNDNLIVENKTBHGGNFQ 414
QY 420 ASAVSISMEXTRIALALIGKLNFTQCTELLNAAMNKGPSCLAEDPSLNHYGKGLDHI 479
Db 415 AAAVNTMEKTRIGLAQIGKLNFTQCTEMNAGMNGRGLPSCLAEDPSLSHYCKGLDIAA 474
QY 480 AAAYSELGHLANPVTTFVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYTSELGHLANPVTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDFKKQDPDLLPTLLQOHLG---TGLDV-NALALVKKALNKLREOTTYYDE 595
Db 535 IDLRALIEFFKKQGPALVSLIDQHFGSAMTGSNLRDELVEKVKTLAKRLQOTSVDLV 594
QY 596 PRMHDAFSYATGTVELLSSSPSANVTLTAVNAKVASAEKALSLTREVRNRFWQTPSQ 655
Db 595 PRMHDAFSYATGTVEVLST---SLSLAANAMKVAABESALSLTROVRETFWSAATS 651
QY 656 APAAHYLSPTRTVLXSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db 652 SPALSYLSPTQILYAFVREELGVARRGDVFLGKQEVITIGSNVSKIYEAIKSGRIHVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 2
US-10-476-198-4
; Sequence 4, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sartasian, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: Hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-476-198-4

Query Match 72.4%; Score 2614; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.2e-194; Indels 14; Gaps 5;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MAPSDLSIATTLANGFTNGSHAAPTSAAGPTSAIRTPGLDGHAAQSLIYQELSD 60
Db 1 MAPSDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQVDIVKMLAA 54
QY 61 PTDDVVELSGSYLTIRDVVGAARKGRVQNDDEIFARVDKSVDFLKAQLQNSVYGTT 120
Db 55 PTDSLLELDGYSILNIGDVVSAARKGRPVRRVQSDSEISKIDKSEVFRSOLNSVYGTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHOLCGVTPTSXSSFSVSGLENTLLPLEVVRAMVIRVN 180
Db 115 GFGGSADTRTEDAISLQKALIEHOLCGVLPSSPFSFLGGLGELNSJPLEVVRAMVIRVN 174
QY 181 SITRGSAVRLVLEALTNFLNHRITPVPKSGISASGDLSPISYAGITGHPDVXK 240
Db 175 SITRGSAVRLVLEALTNFLNHRITPVPKSGISASGDLSPISYAGITGHPDVXK 234
QY 241 VHEGTEKIMFARBAISLFGLEAVVLGKKEGLGVNGTAVSASMATLSLDSHMLSLIQ 300
Db 235 VHEGTEKILVAREMALFNLEPVVLGKKEGLGVNGTAVSASMATLSLDSHMLSLIQ 294
QY 301 ALTALTVEAMVGQGSFAFTHDVCRPHPGQVEVARNIRTLSSGSSFAVEHEEVKVKD 360

Db 295 SLTMTVEANVAGHAGSPHPLHDVTRPHPTQGLEVAQNIKLLGGSFAVAHEEVKVKD 354
QY 361 EGIILRODRPLRTSPQWLGPLVEDMHAYSTLSLE-NNTTNDPLLDVENKQTAAGNFQ 419
Db 355 EGIILRODRPLRTSPQWLGPLVEDLIHHAVALTIEAGOSTTNDNLIVENKTBHGGNFQ 414
QY 420 ASAVSISMEXTRIALALIGKLNFTQCTELLNAAMNKGPSCLAEDPSLNHYGKGLDHI 479
Db 415 AAAYNTMEKTRIGLAQIGKLNFTQCTEMNAGMNGRGLPSCLAEDPSLSHYCKGLDIAA 474
QY 480 AAAYSELGHLANPVTTFVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYTSELGHLANPVTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDFKKQDPDLLPTLLQOHLG---TGLDV-NALALVKKALNKLREOTTYYDE 595
Db 535 IDLRALIEFFKKQGPALVSLIDQHFGSAMTGSNLRDELVEKVKTLAKRLQOTSVDLV 594
QY 596 PRMHDAFSYATGTVELLSSSPSANVTLTAVNAKVASAEKALSLTREVRNRFWQTPSQ 655
Db 595 PRMHDAFSYATGTVEVLST---SLSLAANAMKVAABESALSLTROVRETFWSAATS 651
QY 656 APAAHYLSPTRTVLXSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db 652 SPALSYLSPTQILYAFVREELGVARRGDVFLGKQEVITIGSNVSKIYEAIKSGRIHVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 3
US-10-476-193-19
; Sequence 19, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sartasian, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-19

Query Match 72.4%; Score 2613; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.8e-194; Indels 14; Gaps 5;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;
QY 1 MAPSDLSIATTLANGFTNGSHAAPTSAAGPTSAIRTPGLDGHAAQSLIYQELSD 60
Db 1 MAPSDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQVDIVKMLAA 54
QY 61 PTDDVVELSGSYLTIRDVVGAARKGRVQNDDEIFARVDKSVDFLKAQLQNSVYGTT 120


```

; OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MSC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-19

Query Match      72.4%; Score 2613; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3,8e-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY      1  MAPSLDSLATTANGFTNGNSHAAPTKSAAPPSALRRPPGLDGHAAHOSQLEIVQELLSD 60
Db       1  MAPSLDSLISHSPFANG-----VASAKQAVNGASTNLVAAGSHLPPTQYQTVQVIVKMLAA 54
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61  PDDVDVELSGYSLTYRDVYGAAKRGRRVQNDDEIRAVKDSYVFLAQLONSTYGT 120
Db       55  PDDSTLELDGYSLNIGDVVAARKRPRVVKDSDEIRSDIKDSVEFLSQLSMSYVGYTT 114
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121  GFQGSADRTEDAVSLQKALIEHQLCGVPTPSXSSFSYGRGLENTLPLEVVGAMVIRVN 180
Db       115  GFQGSADRTEDDAISLQKALLHQLCGVLPSPFDSFRLGRGLENLPLEVVGAMVIRVN 174
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      181  SLTRGSAVRLVYVLEALINFLNHRITPVPLEGSLSAGDLSPLSYLIGALTGHDVKYN 240
Db       175  SLTRGSAVRLVYVLEALINFLNHRITPVPLEGLTISAGDLSPLSYIAALISGHDVKYN 234
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241  VLHEGTEKIMPAFEALISGLFAVVLGPHEGLVNGTAVSASMTLSLSDSHMSLSLQ 300
Db       235  VNHGEGEKLTYAREMALFNLPEVVLGPHEGLVNGTAVSASMTLALHDAHMSLSLQ 294
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      301  ALTALTEAVMGQQGSFADFHDVCRPHGQVEVARNITLLSGSSFAVEHEEVKVKXD 360
Db       295  SLTAMTEAVMGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLEGSRFVHHEEVKVKXD 354
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361  EGLLRDRPLRTSPQFGLVEDMMHASTLSL--NNTTNDPLLDVENKQTAHGNGFQ 419
Db       355  EGLLRDRPLRTSPQFGLVEDMMHASTLSL--NNTTNDPLLDVENKTSHHGNGFQ 414
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      420  ASAVSISMKRTLALIGKLNFTQCTELNLNANNGRLPSCLAEDPSLNYKGLDHI 479
Db       415  AAVALNTMKRTLGLAQIGKLNFTQCTELNLNANNGRLPSCLAEDPSLSYCKGLDIAA 474
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      480  AAYASLGLHLNPVYTFQAPMGQAVNSLILSARPTAEANDVLISLASHLYCTQA 539
Db       475  AATYSLGLHLNPVYTHVQAPMANQAVNSLILSARPTESNDVLISLASHLYCTQA 534
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      540  VDIRAELEDFKQDFPLEPTLLQOHG--TGLDV-NALALEVKKALNKRLLEQTTYYLE 595
Db       535  IDIRAELEFFKKQDFGALVSLIDQHFSGMGSNLRDELVEKNNTLAKRLLEQTSYYLV 594
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      596  PRMHDAFSYATGTVVELLSSPSANVLTAVNAKVAASEKALISITREYRNRFQPTSSQ 655
Db       595  PRMHDAFSFAAGTVVEVLST--SLSLAANVAMVYAAASEALISITROVRETFWMAAST 651
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      656  APAAHVALSRTVLYSFVREELGVQARQDVGVQOERTIGSNVRIYEALIDGRINVL 715
Db       652  SPLSLTSLSTRQLLAFVAREELGVKARQDVFLGQOVRTIGSNVSKITFALISGRINVL 711
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      716  VKMLA 720
Db       712  LKMLA 716

RESULT 5
US-10-476-193-24
; GENERAL INFORMATION:
; SEQUENCE 24, Application US/10476193
; APPLICANT: Sarisiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para

```

FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/476,193
CURRENT FILING DATE: 2003-10-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 24
LENGTH: 716
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-193-24

Query Match 72.4%; Score 2612; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 4,66-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATLTANGFTNGSHAPTKSAGPTSAARTRPGLDGHAHQSLQLEIYQELSD 60
DB 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKXLA 54
QY 61 PTDDVVELSGYSLTVRDVVGAAKGRVAVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTSTLELDGYSNLNDGVVSAARKGRPVAVKDSDEIRSKIDKSVFLRSQLSMSVYGVTT 114
QY 121 GFGSGADRTEDAVSLQKALIEHQLCGVTPTXSSEFVGRLNLTPLLEVVRGAMTVRN 180
DB 115 GFGSGADRTEDAVSLQKALIEHQLCGVTPTXSSEFVGRLNLTPLLEVVRGAMTVRN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGATGHPDVVH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGATGHPDVVH 234
QY 241 VHGTEKIMFARBAISLFGLEAVVLGPKGGLGVNGTAVASAMATISLHDSHMLSLISQ 300
DB 235 VHGTEKIMFARBAISLFGLEAVVLGPKGGLGVNGTAVASAMATISLHDSHMLSLISQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNITRLSSGSPFAVEHEEVKXKD 360
DB 295 SLTAMTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNITRLSSGSPFAVEHEEVKXKD 354
QY 361 EGILRODRYPLRTSPQWLGPIVDEDMKAYSTLSLE-NNTTDPNPLDVENKQTAHGNFQ 419
DB 355 EGILRODRYPLRTSPQWLGPIVDEDMKAYSTLSLE-NNTTDPNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISMERTALALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIA 479
DB 415 ASAVSISMERTALALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIA 474
QY 480 AAYASELGLHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLO 539
DB 475 AAYASELGLHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLO 534
QY 540 VDIRAMELDFKQOPDLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
DB 535 IDIRATEFEFKQOPDLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 594
QY 596 PRHMDAFSVATGTVEVELLSSSPANVTLLAVANAKVYASAKAISLTREYARFWQTPSSQ 655
DB 595 PRHMDAFSVATGTVEVELLSSSPANVTLLAVANAKVYASAKAISLTREYARFWQTPSSQ 651
QY 656 APAAAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHL 715
DB 652 SPALSYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHL 711
QY 716 VKMLA 720
DB 712 VKMLA 716

RESULT 6
US-10-476-198-24
; Sequence 24, Application US/10476198

GENERAL INFORMATION:
APPLICANT: Sartasiani, Sima
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the production of Tyrosine, Cinnamic Acid and Para-
TITLE OF INVENTION: Hydroxycinnamic Acid
FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/476,198
CURRENT FILING DATE: 2003-10-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 24
LENGTH: 716
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-198-24

Query Match 72.4%; Score 2612; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 4,66-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATLTANGFTNGSHAPTKSAGPTSAARTRPGLDGHAHQSLQLEIYQELSD 60
DB 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKXLA 54
QY 61 PTDDVVELSGYSLTVRDVVGAAKGRVAVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTSTLELDGYSNLNDGVVSAARKGRPVAVKDSDEIRSKIDKSVFLRSQLSMSVYGVTT 114
QY 121 GFGSGADRTEDAVSLQKALIEHQLCGVTPTXSSEFVGRLNLTPLLEVVRGAMTVRN 180
DB 115 GFGSGADRTEDAVSLQKALIEHQLCGVTPTXSSEFVGRLNLTPLLEVVRGAMTVRN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGATGHPDVVH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGATGHPDVVH 234
QY 241 VHGTEKIMFARBAISLFGLEAVVLGPKGGLGVNGTAVASAMATISLHDSHMLSLISQ 300
DB 235 VHGTEKIMFARBAISLFGLEAVVLGPKGGLGVNGTAVASAMATISLHDSHMLSLISQ 294
QY 301 ALTALTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNITRLSSGSPFAVEHEEVKXKD 360
DB 295 SLTAMTVEAMVGOQGSFAPFIHDVCRPHGQVEVARNITRLSSGSPFAVEHEEVKXKD 354
QY 361 EGILRODRYPLRTSPQWLGPIVDEDMKAYSTLSLE-NNTTDPNPLDVENKQTAHGNFQ 419
DB 355 EGILRODRYPLRTSPQWLGPIVDEDMKAYSTLSLE-NNTTDPNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISMERTALALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIA 479
DB 415 ASAVSISMERTALALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIA 474
QY 480 AAYASELGLHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLO 539
DB 475 AAYASELGLHANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLO 534
QY 540 VDIRAMELDFKQOPDLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
DB 535 IDIRATEFEFKQOPDLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 594
QY 596 PRHMDAFSVATGTVEVELLSSSPANVTLLAVANAKVYASAKAISLTREYARFWQTPSSQ 655
DB 595 PRHMDAFSVATGTVEVELLSSSPANVTLLAVANAKVYASAKAISLTREYARFWQTPSSQ 651
QY 656 APAAAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHL 715
DB 652 SPALSYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHL 711
QY 716 VKMLA 720
DB 712 VKMLA 716

[illegible]

Db	235	VHEGKEKILYAREBAMLFLEPVVLGPKEGLGVNGTVASAMATLALHDAMLSLSQ	294	
Qy	301	ALTAITVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRITLLSGSSFAVEHEEVKXKD	360	
Db	295	SITAMTVEAMVGAGSFPFLHDVTRPHPTQIEVAGNIRKLEGSRFVAHHEEVKXKD	354	
Qy	361	EGILRODRYPLRTSPQMLGVLVSDLIHAHVLITTEGOSTTNPPLDVENKTSHHGNFQ	419	
Db	355	EGILRODRYPLRTSPQMLGVLVSDLIHAHVLITTEGOSTTNPPLDVENKTSHHGNFQ	414	
Qy	420	ASAVSISMEKTRIALALIGKLNFTQCTELNMAANNRGLPSCLAEDPSLNYHGKGLDIHI	479	
Db	415	AAAVANTMEKTRIGLAQIGKLNFTQCTELNMAANNRGLPSCLAEDPSLNYHGKGLDIHI	474	
Qy	480	AAVASELGHLANPYTTVPQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA	539	
Db	475	AAVASELGHLANPYTTVPQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA	534	
Qy	540	VDLRAMEIDFKQFPDPLPTLLQCHLG---TGLDV-NALALEVKAANKRLBQTTDYDE	595	
Db	535	IDLRAXEPEFKQFPALVSLIDQFGSAMTGSNLRDELVEKVKTKLAKRLBQTNISYDLV	594	
Qy	596	PRMHDASFATGTVELLSSSPSANVTTLTAVNAKVASAEKALISLTREVRNRFWQTPSSQ	655	
Db	595	PRMHDASFATGTVELLSSSPSANVTTLTAVNAKVASAEKALISLTREVRNRFWQTPSSQ	651	
Qy	656	APAAHYLSPTRTVLVSFPREEIGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL	715	
Db	652	SPALSYSPRTQILYAFREELGVKARRGDVFLGKQEVTTIGSNVSKIYEAIKSGRIHVL	711	
Qy	716	VKMLA 720		
Db	712	LKMLA 716		
RESULT 9				
US-10-476-193-20				
Sequence 20, Application US/10476193				
GENERAL INFORMATION:				
APPLICANT: Satiastani, Sima				
APPLICANT: Tang, Xiao-Song				
APPLICANT: Qi, Wei Wei				
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-				
FILE REFERENCE: CL1777				
CURRENT APPLICATION NUMBER: US/10/476,193				
CURRENT FILING DATE: 2003-10-28				
NUMBER OF SEQ ID NOS: 24				
SOFTWARE: Microsoft Office 97				
SEQ ID NO 20				
LENGTH: 716				
TYPE: PRT				
ORGANISM: Rhodotorula glutinis				
FEATURE:				
NAME/KEY: MISC_FEATURE				
LOCATION: (181)..(181)				
OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly				
FEATURE:				
NAME/KEY: MISC_FEATURE				
LOCATION: (235)..(235)				
OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro				
FEATURE:				
NAME/KEY: MISC_FEATURE				
LOCATION: (540)..(540)				
OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly				
US-10-476-193-20				
Query Match				
Best Local Similarity				
72.1%; Score 2603; DB 6; Length 716;				
71.7%; Pred. No. 2.3e-193;				
Matches 520; Conservative 85; Mismatches .106; Indels 14; Gaps 5;				
Qy	1	MAPSIDSLATTLANGFTNGSHAAPTYSAGPTSAIRRTPTGIDGHAHQOLEIVQELISD	60	
Db	1	MAPSIDSLISHSPANG-----VASAKQAVNGASTNLVAGSHLPTTQVTDIVERKMLAA	54	
Qy	61	PIDDVVEISGVSLITVDVVGAAKGRARYVNDDEIPARVYSKDFKACLOQNSVYGVTT	120	
Db	55	PIDSTIELDQVSLNIGDVVSAKGRPYRVKQDSIRSKIKXVEFLRSQNSVYGVTT	114	
Qy	121	GGGSGADRTEDAVSLQALIEHQLCVTPTSXSSFSVGRGLENTDPLEVVRGAMVIRVN	180	
Db	115	GGGSGADRTEDAVSLQALIEHQLCVTPTSXSSFSVGRGLENTDPLEVVRGAMVIRVN	174	
Qy	181	SLTRGHSARLVVLEALTNFLNHRITPIVPLRGSIASGDLSPSTYACATIGHPDYKVN	240	
Db	175	SLTRGHSARLVVLEALTNFLNHRITPIVPLRGSIASGDLSPSTYACATIGHPDYKVN	234	
Qy	241	VHEGTEKIMPARAISTFGLAAYVLGPKEGLGVNGTVASASVATLILDSHMLSLSQ	300	
Db	235	XVHEGTEKILYAREBAMLFLEPVVLGPKEGLGVNGTVASASVATLILDSHMLSLSQ	294	
Qy	301	ALTAITVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRITLLSGSSFAVEHEEVKXKD	360	
Db	295	SITAMTVEAMVGAGSFPFLHDVTRPHPTQIEVAGNIRKLEGSRFVAHHEEVKXKD	354	
Qy	361	EGILRODRYPLRTSPQMLGVLVSDLIHAHVLITTEGOSTTNPPLDVENKTSHHGNFQ	419	
Db	355	EGILRODRYPLRTSPQMLGVLVSDLIHAHVLITTEGOSTTNPPLDVENKTSHHGNFQ	414	
Qy	420	ASAVSISMEKTRIALALIGKLNFTQCTELNMAANNRGLPSCLAEDPSLNYHGKGLDIHI	479	
Db	415	AAAVANTMEKTRIGLAQIGKLNFTQCTELNMAANNRGLPSCLAEDPSLNYHGKGLDIHI	474	
Qy	480	AAVASELGHLANPYTTVPQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA	539	
Db	475	AAVASELGHLANPYTTVPQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA	534	
Qy	540	VDLRAMEIDFKQFPDPLPTLLQCHLG---TGLDV-NALALEVKAANKRLBQTTDYDE	595	
Db	535	IDLRAXEPEFKQFPALVSLIDQFGSAMTGSNLRDELVEKVKTKLAKRLBQTNISYDLV	594	
Qy	596	PRMHDASFATGTVELLSSSPSANVTTLTAVNAKVASAEKALISLTREVRNRFWQTPSSQ	655	
Db	595	PRMHDASFATGTVELLSSSPSANVTTLTAVNAKVASAEKALISLTREVRNRFWQTPSSQ	651	
Qy	656	APAAHYLSPTRTVLVSFPREEIGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL	715	
Db	652	SPALSYSPRTQILYAFREELGVKARRGDVFLGKQEVTTIGSNVSKIYEAIKSGRIHVL	711	
Qy	716	VKMLA 720		
Db	712	LKMLA 716		
RESULT 10				
US-10-476-193-21				
Sequence 21, Application US/10476193				
GENERAL INFORMATION:				
APPLICANT: Satiastani, Sima				
APPLICANT: Tang, Xiao-Song				
APPLICANT: Qi, Wei Wei				
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-				
FILE REFERENCE: CL1777				
CURRENT APPLICATION NUMBER: US/10/476,193				
CURRENT FILING DATE: 2003-10-28				
NUMBER OF SEQ ID NOS: 24				
SOFTWARE: Microsoft Office 97				
SEQ ID NO 21				
LENGTH: 716				
TYPE: PRT				
ORGANISM: Rhodotorula glutinis				

```

FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (149)..(149)
/ OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (202)..(202)
/ OTHER INFORMATION: X=Val, Met, Leu, Cys
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (540)..(540)
/ OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (540)..(540)
/ OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-21

```

```

Query Match      72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2.3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

```

```

QY 1 MAPSLDLSLTTLANGTNGSHAPTKSAAGPTSAARTRPGDGHAAHSGLEIVQELISD 60
DB 1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQYQDIVERKMLAA 54
QY 61 PTDVVELSGYSLTVRDVVGAAKRGRRVQNDDEIRARVDKSDVDFLKAQONSYYGVT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRPRVVKSDDEIRKIDKSVFPLRSQLSMVGVT 114
QY 121 GFGGSADRTEDAVSIQKALIHQICGVTPTSXSFSVGRGLENLPLEVVRGAMVIRVN 180
DB 115 GFGGSADRTEDAVSIQKALIEHQICGVLPSFDFRGLGLENLPLEVVRGAMVIRVN 174
QY 181 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
DB 175 SLTRGSAARLVVLEALTNFLNHGITPIVPLRGITISAGDSLPSYIAAISGHPDSKVH 234
QY 241 VLHEGTEKIMFAEALISFGLAEAVVLGPKEGIGLVNGTAVASNAVTLSDHSHMLSLISQ 300
DB 235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGIGLVNGTAVASNAVTLSDHSHMLSLISQ 294
QY 301 ALTALVVEAMVQOQSFARFIDVCRPHGQVEVARNRITLLSGSFVHEEVEVKYKD 360
DB 295 SLTAMTVAMVAGAGSFHFLDVTFRPHPTQIEVAGNIRKLLGSRFVHHEEVEVKYKD 354
QY 361 EGI LRDRYPLRTSPQFLGPIVEDMMHAYSTLSLE-NTTTDNPILLDVENKQTAHGNFQ 419
DB 355 EGI LRDRYPLRTSPQWLGPIVSDLIHAAYVLTIEAGSTTNPILLDVENKTSHHGNFQ 414
QY 420 ASAVSISMEXTRLALALIGKLNFTQCTELLNPAANRGPLSCIAADPBLNTHGGLDHI 479
DB 415 AAAVANTMEKTRLGIAQIKLNFTQCTELLNPAANRGPLSCIAADPBLNTHGGLDIAA 474
QY 480 AAASVSELGHANVTTFVCPAEMNGAVNSLALISARPTAEANDVLSLLASHLYCTIOA 539
DB 475 AATTSSELGHANVTTHVQPAEMANOAVNSLALISARPTESNDVLSLLATHLYCTIOA 534
QY 540 VDLRAMELDPKQFDLPLTLLQOHLG--TGLDV-NALALEVKKALNKRLEQTTVYLE 595
DB 535 IDLRAXEFEPKQFGFAIVSLIDHFGSAMTGSNLRDELVEKVNLTAKRLEQNTSYDLV 594
QY 596 PRMHDAFSYATGVVVLSSSPSANTVLTAVNAKMYASAFAKALISTREVRNRFQTPESQ 655
DB 595 PRMHDAFSYATGVVVLSSSPSANTVLTAVNAKMYAAAESALISTRYVRETFMSAATS 651
QY 656 APHAAYSPRTYLVFVFEELGVQARQDVRVGVQCEFTIGSNVRIEATIKDGINVL 715
DB 652 SPLASTISPTQILYAFVEELGVKARQDVLGQCEVITIGSNVSKITIKSGINNVL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

```

```

RESULT 11
US-10-476-198-20
/ Sequence 20, Application US/10476198
/ GENERAL INFORMATION:
/ APPLICANT: Sarsiani, Sina
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ FILE REFERENCE: CL177
/ CURRENT FILING DATE: 2003-10-24
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 20
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Rhodotorula glutinis
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (181)..(181)
/ OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (235)..(235)
/ OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (540)..(540)
/ OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (540)..(540)
/ OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-20

```

```

Query Match      72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2.3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

```

```

QY 1 MAPSLDLSLTTLANGTNGSHAPTKSAAGPTSAARTRPGDGHAAHSGLEIVQELISD 60
DB 1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQYQDIVERKMLAA 54
QY 61 PTDVVELSGYSLTVRDVVGAAKRGRRVQNDDEIRARVDKSDVDFLKAQONSYYGVT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRPRVVKSDDEIRKIDKSVFPLRSQLSMVGVT 114
QY 121 GFGGSADRTEDAVSIQKALIHQICGVTPTSXSFSVGRGLENLPLEVVRGAMVIRVN 180
DB 115 GFGGSADRTEDAVSIQKALIEHQICGVLPSFDFRGLGLENLPLEVVRGAMVIRVN 174
QY 181 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSIASGDSLPSYIAGATTGHPDVVH 240
DB 175 SLTRGSAARLVVLEALTNFLNHGITPIVPLRGITISAGDSLPSYIAAISGHPDSKVH 234
QY 241 VLHEGTEKIMFAEALISFGLAEAVVLGPKEGIGLVNGTAVASNAVTLSDHSHMLSLISQ 300
DB 235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGIGLVNGTAVASNAVTLSDHSHMLSLISQ 294
QY 301 ALTALVVEAMVQOQSFARFIDVCRPHGQVEVARNRITLLSGSFVHEEVEVKYKD 360
DB 295 SLTAMTVAMVAGAGSFHFLDVTFRPHPTQIEVAGNIRKLLGSRFVHHEEVEVKYKD 354
QY 361 EGI LRDRYPLRTSPQFLGPIVEDMMHAYSTLSLE-NTTTDNPILLDVENKQTAHGNFQ 419
DB 355 EGI LRDRYPLRTSPQWLGPIVSDLIHAAYVLTIEAGSTTNPILLDVENKTSHHGNFQ 414
QY 420 ASAVSISMEXTRLALALIGKLNFTQCTELLNPAANRGPLSCIAADPBLNTHGGLDHI 479
DB 415 AAAVANTMEKTRLGIAQIKLNFTQCTELLNPAANRGPLSCIAADPBLNTHGGLDIAA 474

```

QY 480 AAYASELGHANPVTTFVQPAEMGNQAVNSLALISARTPAENDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHANPVTTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDPFKQOPDLLPTLLQOHLG---TGLDV-NALALEVKKALNKRLEQTTTYLE 595
DB 535 IDLRAXEFPFKQOPGPAIVSLIDQHFSGAMTGSNLDLVEKNKTLANGLEQTNVYDLV 594
QY 596 PRMHDAFSYATGIVVELLSSSPSANYTLTAANAKVAESAELSLTREVRNRFWQTPSSQ 655
DB 595 PRMHDAFSYATGIVVELLSSSPSANYTLTAANAKVAESAELSLTREVRNRFWQTPSSQ 651
QY 656 APAAVLSPTRTVLYSVREELGVQARRGVFVQOQETIGSNVSRITYAIDGGINVYL 715
DB 652 SPALSYLSPRTQILYAFVREELGVARRGVDFLGQEVTTIGSNVSKTYEALKSRINVL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 12
US-10-476-198-21
; Sequence 21, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,198
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X=Val, Met, Leu, Cys
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-21

Query Match 72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2,3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MAPSLDSLATTLANGSTNGSHAAPTYSAGPPTSAIRPTGLDGHAAHQSQLEIYQELISD 60
DB 1 MAPSLDSLISHSANG-----VASAKQAVNGASTNLAAVAGSHLPTTQVTVDIYERKMLAA 54
QY 61 PTDDVVELSGYSTLVEDVGAARKGRVAVQNDDEIRAVDKSVDFLKAQOLNSVYGVTT 120
DB 55 PTSTSTLELDGYSLNLDGVSAARKGRPVAVKSDERIKSDSVFELSQLSMVGVT 114
QY 121 GFGSGADRTEDAVSLQKALIEHQLCGVPTTSXSFVSGRGLENTLPLEVVRGAMVIRVN 180
DB 115 GFGSGADRTEDAVSLQKALIEHQLCGVLPSSFDYFRLGRGLENSLPLEVVRGAMVIRVN 174

QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDISPLSYAGAITGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDISPLSYIAAASGHPDBKHV 234
QY 241 VHEGTEKIMEARREALISFGLEAVYLGRKEGGLVNGVAVASAMATSLDHSMLSLSSQ 300
DB 235 VHEGTEKIMEARREALISFGLEAVYLGRKEGGLVNGVAVASAMATSLDHSMLSLSSQ 294
QY 301 ALTALTVAMVQOQSFAPFIDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
DB 295 SLTAMTVAMVQOQSFAPFIDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 354
QY 361 EGLLRDRYPLFTSQPFLGVEDMMAHXYLSLE-NTTIDNLLDVENKQTAHGNFQ 419
DB 355 EGLLRDRYPLFTSQPFLGVEDMMAHXYLSLE-NTTIDNLLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTLALALIGKLNFTQCTELNNAANRGIPSCIAAEDPSLNYHGKGLDIH 479
DB 415 AAVANTMEKTRLGAQIGKLNFTQCTELNNAANRGIPSCIAAEDPSLNYHGKGLDIH 474
QY 480 AAYASELGHANPVTTFVQPAEMGNQAVNSLALISARTPAENDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHANPVTTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAMELDPFKQOPDLLPTLLQOHLG---TGLDV-NALALEVKKALNKRLEQTTTYLE 595
DB 535 IDLRAXEFPFKQOPGPAIVSLIDQHFSGAMTGSNLDLVEKNKTLAKRLEQTNVYDLV 594
QY 596 PRMHDAFSYATGIVVELLSSSPSANYTLTAANAKVAESAELSLTREVRNRFWQTPSSQ 655
DB 595 PRMHDAFSYATGIVVELLSSSPSANYTLTAANAKVAESAELSLTREVRNRFWQTPSSQ 651
QY 656 APAAVLSPTRTVLYSVREELGVQARRGVFVQOQETIGSNVSRITYAIDGGINVYL 715
DB 652 SPALSYLSPRTQILYAFVREELGVARRGVDFLGQEVTTIGSNVSKTYEALKSRINVL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 13
US-10-476-193-18
; Sequence 18, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,193
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X=Gly, Ala, Ser, Thr
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-18

Query Match 72.0%; Score 2599; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 4,76-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLANGFTNGSHAAPKSAAGPTSAARRTPGLDGHAAHQSOLEIYQELISD 60
Db 1 MAPSLDLSISHPANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYVDIYEKLLAA 54

QY 61 PDDVVELSGYSLIVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQIQNSVGYTT 120
Db 55 PDDSTLELDGYSILNGDVSAARKGRPVKVDSDERKSKIDKSVDFRSLQMSVGYTT 114

QY 121 GFGSADRTTEDAVSLQKALLIEHQCGVPTXSFSFSGRGLENTLPLEVVGAMVIRVN 180
Db 115 GFGSADRTTEKALISQKALLIEHQCGVLPSSFDSPFRGRGLENSLPLEVVGAMVIRVN 174

QY 181 SLTRGSAVRLVLEALNFINHRITPIVPLRGSIISAGDLSPLSYIAGAITGHPDKVH 240
Db 175 SLTRGSAVRLVLEALNFINHGITPIVPLRGITISAGDLSPLSYIAAISGHPDKVH 234

QY 241 VHEGHEKIMPARBAISLFGLEAVVLGPKESGLVNGTAVASAMATLISHSHLISLSQ 300
Db 235 VHEGHEKILLYAREMAALFNLEPVVLGPKESGLVNGTAVASAMATLISHHNLISLSQ 294

QY 301 ALTALTVEMVQOQGSFAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVEVKYKD 360
Db 295 SLTAMTVEMVGHAGSFHPEFLHDVTRPHPTQIEVAGNIRKLLEGSRAVHHEEVEVKYKD 354

QY 361 EGLRQDRPLRTSPQFGLPYEDMMAHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 419
Db 355 EGLRQDRPLRTSPQWGLPVSDDIHAHVAVLTIAAGOSTDNDPLIDVENKTSHHGNFQ 414

QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCIAEDPSLNYHGGDLIHI 479
Db 415 AAAVANTMEKTRIGLAQIOGKLNFTQCTELNNAAMRGPLSCIAEDPSLNYHGGDLIAA 474

QY 480 AAVASLGLHANPVTTFVQPAEMGQAVNSLALISARRTAENDVLSLLASHLYCTLQA 539
Db 475 AAVTSELGLHANPVTTFVQPAEMANOAVNSLALISARRTTESNDVLSLLATHLYCYLQA 534

QY 540 VDIRAMELDFKKQFDPPLPTLIQOHLG---TGLDV-NALALEVKKALKLEQTTTYDLE 595
Db 535 IDIRAXEPEFKQFQPAIYSLIDQHFSGAMTGSNIRDELVEKNKTLAKREQNSIDLV 594

QY 596 PRMHDAFSYATGTVELLSSSPSANVTLTAVNAMKVASAEKALISLREVRNRFMQTPSSQ 655
Db 595 PRMHDAFSYATGTVEVLST---SLSLAAVNAAMKVAASAEKALISLREVRTPFMSAATS 651

QY 656 APRAHAYLSRTVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHL 715
Db 652 SPALSTYLSPTQILYAFVREELGVQARRGDVFLGQEVLTIGSNVSKIYEAIKSGRINNL 711

QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 14
US-10-476-198-18
; Sequence 18, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariaatani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18

LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X=Gly, Ala, Ser, Thr
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-18

Query Match 72.0%; Score 2599; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 4,76-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLANGFTNGSHAAPKSAAGPTSAARRTPGLDGHAAHQSOLEIYQELISD 60
Db 1 MAPSLDLSISHPANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYVDIYEKLLAA 54

QY 61 PDDVVELSGYSLIVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQIQNSVGYTT 120
Db 55 PDDSTLELDGYSILNGDVSAARKGRPVKVDSDERKSKIDKSVDFRSLQMSVGYTT 114

QY 121 GFGSADRTTEDAVSLQKALLIEHQCGVPTXSFSFSGRGLENTLPLEVVGAMVIRVN 180
Db 115 GFGSADRTTEKALISQKALLIEHQCGVLPSSFDSPFRGRGLENSLPLEVVGAMVIRVN 174

QY 181 SLTRGSAVRLVLEALNFINHRITPIVPLRGSIISAGDLSPLSYIAGAITGHPDKVH 240
Db 175 SLTRGSAVRLVLEALNFINHGITPIVPLRGITISAGDLSPLSYIAAISGHPDKVH 234

QY 241 VHEGHEKIMPARBAISLFGLEAVVLGPKESGLVNGTAVASAMATLISHSHLISLSQ 300
Db 235 VHEGHEKILLYAREMAALFNLEPVVLGPKESGLVNGTAVASAMATLISHHNLISLSQ 294

QY 301 ALTALTVEMVQOQGSFAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVEVKYKD 360
Db 295 SLTAMTVEMVGHAGSFHPEFLHDVTRPHPTQIEVAGNIRKLLEGSRAVHHEEVEVKYKD 354

QY 361 EGLRQDRPLRTSPQFGLPYEDMMAHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 419
Db 355 EGLRQDRPLRTSPQWGLPVSDDIHAHVAVLTIAAGOSTDNDPLIDVENKTSHHGNFQ 414

QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNNAAMRGPLSCIAEDPSLNYHGGDLIHI 479
Db 415 AAAVANTMEKTRIGLAQIOGKLNFTQCTELNNAAMRGPLSCIAEDPSLNYHGGDLIAA 474

QY 480 AAVASLGLHANPVTTFVQPAEMGQAVNSLALISARRTAENDVLSLLASHLYCTLQA 539
Db 475 AAVTSELGLHANPVTTFVQPAEMANOAVNSLALISARRTTESNDVLSLLATHLYCYLQA 534

QY 540 VDIRAMELDFKKQFDPPLPTLIQOHLG---TGLDV-NALALEVKKALKLEQTTTYDLE 595
Db 535 IDIRAXEPEFKQFQPAIYSLIDQHFSGAMTGSNIRDELVEKNKTLAKREQNSIDLV 594

QY 596 PRMHDAFSYATGTVELLSSSPSANVTLTAVNAMKVASAEKALISLREVRNRFMQTPSSQ 655
Db 595 PRMHDAFSYATGTVEVLST---SLSLAAVNAAMKVAASAEKALISLREVRTPFMSAATS 651

QY 656 APRAHAYLSRTVLYSFVREELGVQARRGDVFGVQOCTIGSNVSRITYEAIKQGRINHL 715
Db 652 SPALSTYLSPTQILYAFVREELGVQARRGDVFLGQEVLTIGSNVSKIYEAIKSGRINNL 711

QY 716 VKMLA 720
Db 712 LKMLA 716

```
RESULT 15
US-10-476-193-17
; Sequence 17, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariafian, Sima
; APPLICANT: Tang, Xiao-song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,193
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X= Gly, Ser, Ala, Thr
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X=Val, Met, Ieu, Cys
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X=Ala, Gly, Ser, Pro
; NAME/KEY: MISC FEATURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: X=Gly, Ala, Ser, Thr, Pro
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-17

Query Match      71.4%; Score 2579; DB 6; Length 716;
Best Local Similarity 71.0%; Pred. No. 1,7e-191;
Matches 515; Conservative 86; Mismatches 110; Indels 14; Gaps 5;

QY      1  MAEPLSLATTLANGFTNSHAAPTSAAGPTSAARTRPGDGHAAHSGLEIVQELISD 60
DB      1  MAEPLSLISHSFANG-----VASAKQAVNGASTNLAAVAGSHLPTTYQTVQVIVKMLAA 54
QY      61  PTDDVVELSGSYLTVDVVGAAKRGRRVQNDDEIRARVDKSVDFLKAQLONSIVGYTT 120
DB      55  PTSTJLELDGYSLNLDGVVSAARKGRPVAVKXDSDEIRSKIDKSVFELRSQLSMVGVT 114
QY      121  GFSGSADTRTEDAVSLQKALIEHQLOGVTPTXSXSFVGRGIENTLPLEVVRGAMVIRVN 180
DB      115  GFSGSADTRTEKXALSLQKALBHXLCGLVPSFDFXFLGRGLENSLPLEVVRGAMTIRVN 174
```

```
QY      181  SLTRGSAVRVLVLEALNPNLNRITPTVPLRGSIASGDISPLSYAGAITGHPDVKH 240
DB      175  SLTRGIXAVRVLVLEALNPNLNXITTPVPLRGTISAGDISPLSYAAISGHPDKVH 234
QY      241  VHEGTEKIMFARERASLFGLEAVVLGPKREGIGLVNGTAVASNAATSLDHSMLSLSSQ 300
DB      235  XHHEGKEKILYREEMALPNLEPVLGPKREGIGLVNGTAVASNAATLALDHAMLSLSSQ 294
QY      301  ALTALTVEAMVQGGSPAFPHIDVCRPHGQVEVARNIRTLSSGSPAVHEEVEVKYKD 360
DB      295  SLTAMTVEAMVGHAGSPFLHDVTRPAPTQIEVAGNIRKLSSGSPAVHHEEVEVKYKD 354
QY      361  EGLRQDRYPLRTPSGPFIQVLEDMMHAYSLSE--NNTTDNPLDVENKQTAHGNFQ 419
DB      355  EGLRQDRYPLRTPSGPFIQVLEDMMHAYSLSE--NNTTDNPLDVENKQTAHGNFQ 414
QY      420  ASAVSISMEKTRIALALIGKLNFTQCTELLNAAANRGIPSCLAEDPSLNVHGKGLDHT 479
DB      415  AAAVANTWEKTRLGIAQIGKLNFTQCTELNAGANNRGLPSCLAEDPSLSYHCKGLDIAA 474
QY      480  AAYASFLGHLANPYTTVPQAPAKNGQAVNSALIASARTAEANDVLSLLASHYCTLOA 539
DB      475  AAYTSLEGLHANPVTVHVPAPAKMANQANSLALISARRTESNDVLSLLATHLYCYLOA 534
QY      540  VDLRAVELDPKKQDPDLPTLLOOHLG---TGLDV--NALALEVKKALNKLEQTTVDLE 595
DB      535  IDLRALIEPFKKQFGPALVSLIDQHFSGAMTGSULREIYENKNTLAKLEQTNISYDV 594
QY      596  PRMHDAFSYATGVTELLSSPSSANVTITAVANAKVASAEKALISLREVENRFRWQTPSSQ 655
DB      595  PRMHDAFSFAAGTVEEVSST---SLSLAAVNAKVAASALSLTRQVETFTWSAASTS 651
QY      656  APAHAAYLSPRTVLYSPFREELGYQARRGDVFGVQOETIGSNVSRIFYEAIKGRINHYL 715
DB      652  SPALSYLSERTQILYAFVREELGYKARRGDVFLGKQEVTTIGSNVSKIFYEAIKGRINNVL 711
QY      716  VKOLA 720
DB      712  LKOLA 716
```

Search completed: September 9, 2004, 10:21:12
Job time : 46.8174 secs

[illegible]

RESULT 2
A29607
Phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodospiridium toruloides) (strain IT
C:Species: Rhodospiridium toruloides
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
C:Accession: A29607
R:Anson, J.G.; Gilbert, H.J.; Oram, J.D.; Minton, N.P.
Gene 58, 189-199, 1987
A:Title: Complete nucleotide sequence of the Rhodospiridium toruloides gene coding for p
A:Reference number: A29607; MUID:88112870; PMID:2828184
A:Accession: A29607
A:Molecule type: DNA
A:Residues: 1-693 <ANS>
A:Cross-references: GB:MI8261; NID:g169745; PIDN:AAA33863.1; PID:g169746
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase, carbon-nitrogen lyase
F:18-130/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:185/Modified site: dehydroalanine (Ser) #status predicted

Query Match	71.3%;	Score 2573;	DB 2;	Length 693;
Best Local Similarity	74.9%;	Pred. No. 6.5e-163;		
Matches 507;	Conservative	83;	Mismatches	79;
			Indels	8;
			Gaps	4;

[illegible]

```

Db      320 VHHHEEVKAVKDDGGILRQDRYPRTSPQWLGEPVNSLIHAHVLTTEAQSTTNDPLIDV 379
QY      408 ENKQTAAGNGFQASAVSISMEKTRLTALAIIGKLNFTQCTELLNAAMNGLPSCIAEDPS 467
Db      380 ENKTSHHGNGFQAAAVANTMEKTRLTGLAQIGKLNFTQCTELLNAAMNGMLPSCIAEDPS 439
QY      468 LNHGKGLDHIHAAYSEGLHANPVTTVQAPAEKMGQAVNSIALISARPTHEADVLSL 527
Db      440 LSYCKGLDIAAAAYSEGLHANPVTTHQAPAEKMANQAVNSIALISARPTESADVLSL 499
QY      528 LLSHVICTLIQAVDLFAMELDFFKQFDPLPFLTLOOHG---TGLDV-NALALEYKALN 583
Db      500 LILATHLYCVQALDLDAIEEFKQKQGPALVSLIDHFGSANTGSNLDELVEKKNKTLA 559
QY      584 KRLBQTTTYDLERWMDAFSYATGIVELLSSPSANVTLLTAVNAKVAASAERKALSITRE 643
Db      560 KRLBQNTSYDLVRWMDAFSPAAGTIVEVLST---SLSLAVNAVMKVAALASALSITRQ 616
QY      644 VNRKFMQTPBSQAPAAAYISPRTRVLYSVRRELVQVQARGVFPVGVQOETIGSVNSKTY 703
Db      617 VRETFVSPASTSSPALSYLSPRTQILYAVRBEELGVKARQGVDFGKQEVATIGSVNSKTY 676
QY      704 EATKDGRIHNVLMKLA 720
Db      677 EATKSGRIHNVLMKLA 693

```

RESULT 3
 S01999
 phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodotorula rubra)
 C:Species: Rhodotorula rubra
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
 C:Accession: S01999
 R:Filpula, D.; Vasler, C.A.; Levy, A.; Sykes, A.; Strausberg, R.L.
 Nucleic Acids Res. 16, 11381, 1988
 A:Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula
 A:Reference number: S01999; PMID:89083582; PMID:2205749
 A:Accession: S01999
 A:Molecule type: DNA
 A:Residues: 1-713 <F5L>
 A:Cross-references: EMBL:X13094; NID:G3284; PIDN:CA31486.1; PID:G9255942
 R:Vasler, C.A.; Strausberg, R.L.; Sykes, A.; Levy, A.; Filpula, D.
 Nucleic Acids Res. 16, 11382, 1988
 A:Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal gen
 A:Reference number: S01990; PMID:89083583; PMID:3205750
 A:Accession: S01990
 A:Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-674 <VAS>
 A:Cross-references: EMBL:X13055
 C:Genetics:
 A:Introns: 110/2; 139/2; 372/3; 428/2; 482/3
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 E:217-219/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 I:218/Modified site: dehydroalanine (Ser) #status predicted

Query Match 69.2%; Score 2499.5; DB 2; Length 713;
Best Local Similarity 69.6%; Pred. No. 5.2e-158;
Matches 51; Conservative 83; Mismatches 105; Indels 35; Gaps 9;

```
QY 1 MAASLUSLTTJLANGTNGSHAAPT-----KSAAGTSLRRPRTGJGHAHAQSOUEI 53
QY 1 MAASVSVSIATSVANSLSNGIHPAAAANGEDVHKKTAGAGSLPTT-----ETTQDIDI 52
Db 1
QY 54 VOELJSDP-TDVEVELSGSYLTFRDVGVAARKRRFRVQUNDDEIIRAVDKSVDELKAQLO 112
QY 53 VERILIDAGATDQIKUDGYTTLFLGDVGVAARGRGFRVKAADSPHIREKIDAEVEFLRTOLD 112
Db 53
QY 113 NSRYGTTTGGGASADIRTEDAVSLQKALJIEHQLGVPTFSXSSFSVGRGLENTPLEVVR 172
QY 113 NSRYGTTTGGGASADIRTEDAVSLQKALJIEHQLGVPTFSXSSFSVGRGLENTPLEVVR 172
Db 113 NSRYGTTTGGGASADIRTEDAVSLQKALJIEHQLGVPTFSXSSFSVGRGLENTPLEVVR 172
```

QY	173	GANVIRVNSITRGHSAVLYVLEALTNELNRIITVYLRSSISASGDSLPSIYAGAIT	232
Db	173	GANITIRVNSITRGHSAVRIVLEALTNELNHOITVYLRQTIISASGDSLPSIYASISIT	232
QY	233	GHPDVAVNHLHEGTEKIMFAREAISLEFGEAVLVEPKFEGELGVNNGTAVASAMATLSHDS	292
Db	233	GHPDSKVHV--DC--KIMSAQRAIALKSGQPVLYGPKGELGVNGTAVASAMATLALTDA	288
QY	293	HMLSLLSQALITATVEAMVGOQGSFAPFIHDVCRPHQGVAVANNIRLLSGSSFAVEHE	352
Db	289	HVLSLLAQALITATVEAMVGHAGSFHPFLHDVTRBHPQIEVANNIRLLGSSKYAVHHE	348
QY	353	BEKVAVDDDEGLIKODPVPPLRTSPQELGVYEDMMHAYSTLSLE--NNTTNDPPLDVENKQ	411
Db	349	TEKVAVDDDEGLIKODPVPPLRCSPOWIGPLVSMIAHAYSLSEAGQSTDPPLDLENKK	408
QY	412	TAHGNSFOASAVASISMEKTRLALILGKINFQCTELNANMRNGLPESCLAAEDPSLNTH	471
Db	409	THHGAFPMASSVGNITMEKTRLALVLMGKYSFQLTMLNAGMNTALBCLAAEDPSLSYH	468
QY	472	GKGLDIIHAAVSELGHLANPVYTFVQAPAMGNQAVNSLALISARTAEANDVLSILLAS	531
Db	469	CKGLDIAAAAYTSELGHLANPVSTHVQAPAMGNQAINSLALISARTAEANDVLSILLAT	528
QY	532	HLIYCTIQAVDLRAMELDFKKQDFDPLPTLLOOHLGTGDVNAALL-----EYKALKREL	586
Db	529	HLICVIOAVDLRAMEEHTRKAFEEPPVTELLKXHPG-----ALATAEYEDVKRSIYKRL	582
QY	587	EORTTYDLERPHMDASVATGVVELLSSPSANTTLTAVANAMKVASAEKAIILSTREVN	646
Db	583	QONNSITDLERPHMDTSVATGAVVELLAQO---EYSLASLNAMKVACAEKAIALTREVD	639
QY	647	RFNQTSSQAFAHAYLSPTRVLYSFVREELGVQARRDGVGVQOETIGSNVARIYEAI	706
Db	640	SFAAASSSS6PALKYLSPTRVLYSFVREEVGVKARRDGVLYGQEVITIGTVNARIYEAI	699
QY	707	KDGRINHVLVKMLA 720	
Db	700	KSGCIAPVLVKMMA 713	

```

RESULT 4
GB84787
Phenylalanine ammonia lyase (PAL) [imported] - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C|Accession: G84787
R|Lili, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.|Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vanken, S.E.; Umayam, L.; Tallon, B.;
Eusse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nucleotide 402 761-768 1999
A|Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A|Reference number: A64420; MUID:20083487; PMID:10617197
A|Accession: G84787
A|Status: Preliminary
A|Molecule type: DNA
A|Residues: 1-725 <STO>
C|Cross-references: GB:AEO02093; NID:g6598547; PIDN:AAD8156.2; GSPDB:GRN00139
C|Genetics:
A|Gene: At2g37040
A|Map position: 2
C|Superfamily: histidine ammonia-lyase

Query Match          28.9%; Score 1041.5; DB 2; Length 725;
Bect Local Similarity 36.5%; Pred. No. 3,7e+61;
Matches 272; Conservative 138; Mismatches 258; Indels 77; Gaps 21;

18 NGSHAAPFKSAAGPISAL-----RRTPGUDGHAAHQ---SOLEIVQELIISD 60
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db    4  NGAH----KSNGGVDAHLCGGDITKKNVYINAEDPLWGAHAEGMKSHLDDEVKRYAE 59
      :|::|:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

61 PTDVVDELGSGLTVRDVYGAAAKRGRRVRVQNDDIEPARVDKSVDFLKAOQLNSV--YGV 118
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

D	b	60	FKRPVYNNLGGELITTTGGVAAISTTGNKXKELSTAAQVNAASDWMMEMNKGTSYGV	119
Q	y	119	TTGFGGSADRTTEDAVSIQKALLIEHOLCGVTPITSXSBSYGRGLENTLPLEVRGANVIR	178
D	b	120	TTGFGATSHRRFRKNQVAIQKELIRFLNAGI-----FGSKTKESHLPSPASATRAAALVR	172
Q	y	179	VNSLIRGSAAVYLVLAELTNFLNHRIPRIYVLRGSIASGSDISPLYIAGATGHEVDK	238
D	b	173	INTLIGRSGIFELLEAITSPFNNTTSPRLPGITTTASGDLVPLSYIAGLTGRNSK	232
Q	y	239	VHVLHGTE-KIMFAREAISLFGLEA--VVLGKEGIGLVNGTAVASMAATLSLHDSHML	295
D	b	233	A-----TGPRGKALTEEEAFKLAGISSGCFDLPQREGIALVNGTGVSGMSMWLFETNVL	288
Q	y	296	SLLSQALTTATYEAANVWGQGSFAPRIHVCSRHFGQVEVARNRTLLSSASPAVEHEEV	355
D	b	289	SVLAELTSLVAPEVNSGR-ETDHLTRLHHFPGQIEAAIYEHILDGSSY-WKLAQKL	346
Q	y	356	KVWDEGILRCQRYELRTSPQFLGVEDMMHAYSTLSLNTTTDNPILDVENKQTANG	415
D	b	347	HEMDPLQKRCQRYALRTSPQWLGQFIVHIVAKTSIEREINSVNDPRLIDVSRNKAIHG	406
Q	y	416	GNTFQASAVISNEKTRLLALIGKINPQCTELLNAAMNGLPCL-AAEDPSLNTYHGK	474
D	b	407	GNTFQGPICVSDNTRLAAIIGKIMFQFSELVNDYNNGLPNTLASNPISLDYFPG	466
Q	y	475	LDIHIAVASELGHIANPVTFVCPAEKMGQAVNSLALISARPAEANDVLSLLASHLY	534
D	b	467	AEIAMAASYSELQYIANPVTHSVGSAEQRNDVNSLGLISRKTSKSEAVDIKLKMTSTFLV	526
Q	y	535	CTLQAVDLRAMELDFKQDPRLPTLLQOHLGTGLDVNALAEVKKALNKL-----EQ	588
D	b	527	ALCOAAVDLHLEENRQTVKNTVSGVAKKVLTTG--VNG-ELHPSRCECKDLLKVYDREQ	583
Q	y	589	TTTTYDLERPMHDAFVYAGTVE-----LLSSPSANVLTLVNMAKVASAEKAI	638
D	b	584	VYTYADP-----CSAITYPLIQKRLQVIVHALINGSEKNATSYTFH-KIGAFEEEL	635
Q	y	639	S--LTREY--NRMQTPSSQAPAHAYLSPTRLVLSFVREELGVQARQDVFVGOQST	694
D	b	636	KAVLPREVEAARAADNGTSALP-NRIKECSYPLRYRVEELGT-----ELLTGEKVT	689
Q	y	695	IGSNVSRITYEAIKQGRINHVLVKYL	719
D	b	690	PGSEEPKVFYALCEKGIIDPMMECI	714

RESULT 5
S52990
Phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 14-Jul-1995 #sequence_rev1995 #ext_change 16-Jul-1999
C/Accession: S52990; PQ0140
C/Manner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant MOL. Biol. 27, 327-338, 1995
A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A/Reference number: S52990; NUID:95195160; PMID:7886622
A/Accession: S52990
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-725 <MAN>
A/Cross-references: EMBL:I33677; NID:G497418; PIDD:AAIC18870.1; PID:G497419
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R/Ohl, S.; Hedrick, S.A.; Chory, J.; Lamb, C.U.
Plant Cell 2, 837-849, 1990
A/Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsis
A/Reference number: PQ0140; NUID:93005677; PMID:2152331
A/Accession: PQ0140
A/Molecule type: DNA
A/Residues: 1-240 <OHL>
A/Cross-references: GB:J62747; NID:G16408; PIDD:CAA44609.1; PID:G16409
A/Experimental source: Ecotype Columbia
A/Comment: Phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to L-

C/Genetics:
A/Gene: PAL1
A/Intons: 143/2
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F:211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:212/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.7%; Score 1035.5; DB 2; Length 725;
Best Local Similarity 36.4%; Pred. No. 9.2e-61;
Matches 271; Conservative 138; Mismatches 259; Indels 77; Gaps 21;

```
QY 18 NGSHAPTKSAGTSL-----RRTGDLGHAHQ---SOLEYQELSD 60
DB 4 NGAH---KSNGGVDMLCGGDIKTKMNVINAEDPLNGLAAEDQKSHLDEVKRYAE 59
QY 61 PTDDVELSGSLTVRDVVGAAKGRVAVQNDDEIRAVKXVDPLKALQNSV--YGV 118
DB 60 FRKRVNLTGGETLTIGVAAISTIGKSVKVELSETARAGVNASDWMVSMKGTDSYGV 119
QY 119 TTGRGSADTRTDANSLOKALIEHQLCGVTPKXSSFSVSGLENTLPLEVKGAMVIR 178
DB 120 TTGGATSHRRTKGVALQKELRFLNAGI-----FGSTKETSHTLPHSATRAMMIVR 172
QY 179 VNSLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYAGATGHPVK 238
DB 173 INTLQSGSIRFEILEITATISFLNNITPSLRGITTSGLVPLSYIAGLIGRPSK 232
QY 239 VHVHBSGE-KIMFAEALSLFGLA--VVLGPKGGLVNGTAVSASMATLSLHDSML 295
DB 233 A---TGPNGEALTLAEAFKLAGISSGFEDLPKKGALVNGTAVGSGMASVPLFEITVL 288
QY 296 SLTSOALTLTVAMVGQGGSPAFPIHDVCRPHGQVEAVARIRTLSSGSPAVEHEEV 355
DB 289 SVLAELISAVFAEVMGKR-EFTDHLTHLKHHPGOIEAAMHEILDGSSY-MKLAOKL 346
QY 356 KVXDDGILRQDRYPLRTSPQFLGLVEDMHAYSTLSLENNTTDNLPLENKKQTAG 415
DB 347 HEMDPLQKPRQDYALRTSPQWLGPOIEVIRATKSIEREINSVNDNPLIDVSRKAHG 406
QY 416 GNGQASAVSISMEKTEFLALIGKLNFTQCTELNAAAMRGPSCL-AAEDPSLNYHKG 474
DB 407 GNGQGTPIGVMKNTLALIRALIGKLPFQSELVNDPYNGLPSMLTTSRNPSLDYGK 466
QY 475 LDHIAAYASELGHANPYTTFVQPAEMGNQAVNSIALISARTAEANDVLSLLASHLY 534
DB 467 ABIAVMSYSGELQYLANPVTSHVQSAEQHNDVNSGLISSRKTSSEAVDILKMTSTFLV 526
QY 535 CTIQAVIDRAMELDFKQDFPLPTLLOQLGTLGDVNALALEVKKALKRLE-----EQ 588
DB 527 AICQAVDLRLLENLQVTKNTVSQVAKKVLTTG--VNG-ELHPSRFCEKDLKRVDRQ 583
QY 589 TTYDLEPRWHDASFATGTAVE-----LSSSPSANTVLTANAKVSAEKA 638
DB 584 VYTYADDP-----CATYPLIQKLRQVIVDHALVNGSEKNAVTSIH-KIGAFBEEL 635
QY 639 S--LTREVR--NRFQTPSSQAPAHAYLSPTFRVLYSYREBELVQARGDVFVQOET 694
DB 636 KAVLPKEVAAARAAYNGTSAIP-NRIKCRSGYPLRYFVREBELG-----ELLTGEKVT 689
QY 695 IGSNVRIRYEALIDGRINHVLKML 719
DB 690 PGEFPKVFYALCEGKIIDPMMECL 714
```

RESULT 6
S48726
phenylalanine ammonia-lyase (EC 4.3.1.5) 3 - parsley
C/Species: Petroselinum crispum (parsley)
C/Date: 18-Oct-1995 #sequence_revision 12-Apr-1996 #text_change 16-Jul-1999
C/Accession: S48726; S56035
R/Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994

A/Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase i:
A/Reference number: S48725; MUID:95010141; PMID:7925477
A/Accession: S48726
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-718 <App1>
A/Cross-references: EMBL:X81159
R/Appert, C.
Submitted to the EMBL Data Library, August 1994
A/Reference number: S56035
A/Accession: S56035
A/Molecule type: mRNA
A/Residues: 1-58, 'E', 60-718 <App2>
A/Cross-references: EMBL:X81159; NID:9535007; PIDD:CAA57057.1; PID:9535008
C/Genetics:
A/Gene: PAL3
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F:204-206/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:205/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.2%; Score 1017.5; DB 2; Length 718;
Best Local Similarity 36.5%; Pred. No. 1.4e-59;
Matches 269; Conservative 128; Mismatches 275; Indels 65; Gaps 19;

```
QY 14 NGFTGSHAPTKSAGTSLRTRP-----GLDGHAAHQSOLEYQELSDPTDDVELS 69
DB 5 NGTTNG-HA---NANGIDLCKKEDPLNMGVAEALITGSHLDVKNVAEYKRPVYKLG 59
QY 70 GXLTVRDVVG-AAKGRVAVQNDDEIRAVKXVDPLKALQNSV--YGVTTGFGSA 126
DB 60 GELTLISQVAAISARDGSGVKELEBARAGVKASDWMVSMKGTDSYGVTTGFGATS 119
QY 127 DTRTDANSLOKALIEHQLCGVTPKXSSFSVSGLE--NTLPLEVKGAMVIRVNSLT 183
DB 120 HRRTKQGGALQKELRFLNAGI-----FGSGAAGNNTLPHSATRAMMIVRINTLL 170
QY 184 RGHSAVRLVLEALTNFLNHRITPIVPLRGSISASGDLSPSYAGATGHPVKVHVA 243
DB 171 QVSGIRFEILEITATISFLNNITPCLPRGITTSGLVPLSYIAGLIGRPSKA---V 227
QY 244 EGTEKIMFAEALSLFGLA--VVLGPKGGLVNGTAVSASMATLSLHDSMLLSQA 301
DB 228 GPTGVTLSPBEAFKLAGVGGFPFLOPKXGALVNGTAVGSGMASVPLFANILAVLAEV 287
QY 302 LFTALTVAMVGQGGSPAFPIHDVCRPHGQVEAVARIRTLSSGSPAVEHEEVKXKDE 361
DB 288 WSAIFAEVMGKR-EFTDHLTHLKHHPGOIEAAMHEILDSAG-VKAAOKLHEMDPL 345
QY 362 GILRODRYPLRTSPQFLGLVEDMHAYSTLSLENNTTDNLPLENKKQTAGNFQAS 421
DB 346 QKPKQDRYALRTSPQWLGPOIEVIRATKSIEREINSVNDNPLIDVSRNAKIGGPFQGS 405
QY 422 AVSISMEKTRFLALIGKLNFTQCTELNAAAMRGPSCLA-AAEDPSLNYHKGDLHIA 480
DB 406 PIGVSMNTRILALIGKLPFQSELVNDPYNGLPSNLSSGRNPSLDVGFQGAETANA 465
QY 481 AYASELGHANPYTTFVQPAEMGNQAVNSIALISARTAEANDVLSLLASHLYCTLOAV 540
DB 466 SYCELOFLANPVTNHVQSAEQHNDVNSGLISSRKTSSEAVELIKLMTSTFLVGLCQAI 525
QY 541 DLRAVELDFKQDFPLPTLLOQLGTLGDVNALALEVKKALKRLEQ-----588
DB 526 DLRLLENLKSTVKNTVSQVAKRVLTWG--VNG-ELHPSRFCEKDLRVVDRREYITAYTD 582
QY 589 ---TTTYDLEPRWHDASFATGTAVVELSSSPSANTVLTANAKVSAEKAIS--LTRE 643
DB 583 DPCSATYPLMQKRE-----TLVEHALNNGKERNLSTISQKIAAFEDDELKALLPRE 635
QY 644 VNN-RFQWTPSSQAPAHAYLSPTFRVLYSYREBELVQARGDVFVQOETIGSNVSR 702
DB 636 VETARAALSGNPAPIPRIKCRSGYPLRYFVREBELG-----EYLTGEKVRSGEEREXV 690
```

QY 703 YEAKDGRINHVLYKML 719
: : : : :
Db 691 FTAMSKGEIIDLLECL 707

RESULT 7

66343 phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco

C/Species: Nicotiana tabacum (common tobacco)
C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999

C/Accession: S66343
R/Fukusawa-Akeda, T.; Kung, S.; Watson, J.C.

Plant Mol. Biol. 30, 711-722, 1996

A/Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicotiana
A/Reference number: S66343; WUID:96194455; PMID:8624404

A/Accession: S66343

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-715 <FUK>

A/Cross-references: EMBL:W84466; NID:g170349; PIDN:AAA34122.1; PID:g170350

C/Genetics:

A/Gene: PAL1

A/Intons: 133/2

C/Superfamily: histidine ammonia-lyase

C/Keywords: ammonia-lyase; carbon-nitrogen lyase

F/201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted

F/202/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.2%; Score 1016.5; DB 2; Length 715;
Best Local Similarity 36.0%; Pred. No. 1.6e-59;

Matches 270; Conservative 130; Mismatches 259; Indels 91; Gaps 19;

QY 13 ANGFTN-GSHAAFTKSAAGPTSAALRTPGLDGHAHQSOLEIVQELSDPTDDVVELSGY 71
: : : : :
Db 3 SNQHVNGGFELCKKADPLNWEAAESLRG----SHLSEVKKQWSEFRKPMVKLGE 57

QY 72 SLTVADVVGAA--RKGRRVYQNDDEIRAVDKSVDFLKAQIONSV--YGVTTGFGGSA 126
: : : : :
Db 58 SLTVAAQVAIAIVRDKSANGVKVLESEARAYKASDWMDSMKNKGTGYGVTTGFGATS 117

QY 127 DFTEDAVSLQKALIEHQLOGVTPTSXSFVGRGLENTLPLEVVGAMVIRVNSLTRGH 186
: : : : :
Db 118 HRTKKGALQKELIRFLNAGV-----FGNGTETSHTLPHSATRAAMLVRINTLQGY 170

QY 187 SAARLVLEALTNFNLHRTITPVLKGSISASGDLSPSYIAGATGHPDKVYVHEGT 246
: : : : :
Db 171 SGIRFELLETITKLINSITPCLPLRGTTASGDVPLSYIAGLTGRPNKSA--VGNG- 228

QY 247 EKIMFARBAISLFGLEA--VVLGKREGGLVNGTAVSASMATLSLHDSHMLSLSQALTA 304
: : : : :
Db 229 -ETLNBEARVAVGVNGFPELQKEGALVNGTAVSGMASMWLFDNITLAVNSEVLSA 287

QY 305 LTVAVVGGQGSFAPFIDVCRPHQGVAVANIRITLSSGSSFAVHEBEVKVYDDEGL 364
: : : : :
Db 288 IFAEVNNKGP-EFTDHLTHKLKHHPQIEAAIMEHILDGSSY--VKAQKLEMDPLQKP 345

QY 365 RORRPLRTSPQFLGPEVDMMAVSTLSLNNTTDNDPLDVEKQOTAGHGFQASAVS 424
: : : : :
Db 346 KQRYALRTSPQFLGQIEVIRATKMIREINSVNDNPLIDVSNKKAHGNFGGTIG 405

QY 425 ISMEKTRIALALIGKINFTQCTELLNAAAMRGIPSL-AAEDPSINHGKGLDIIAAYA 483
: : : : :
Db 406 VSMNDARALASIGKIMFAQFSELVNDYNNGLPSNLTASRNPISIDYFGKGAELIMASYC 465

QY 484 SELCHLANPTTVTPVQEMKNOAVNSIALISARTAEANDVLSLLASHYTTQAVDLR 543
: : : : :
Db 466 SELQFLNPVTHVQSAEHNQDVNSLGLISAKRTAEAVDILKLMSSTYLVALCOATILR 525

QY 544 AMELDFKKQDPLLPFTLLQOHLGTGLDVNALALEVKAANKRLQOTTYDLEFRWHA-- 601
: : : : :
Db 526 HLEENLK-----NA-----VKNIVSQVAKKTLTMGANGELHPARF 560

QY 602 -----FSYATG-----TVVELSSSPSANTVTLTAVNAKVAS 633
: : : : :
Db 602 -----FSYATG-----TVVELSSSPSANTVTLTAVNAKVAS 633

Db 561 CEKELLRTVDBREYFAVADDPSCNYPIMQKLRQVLDHAMNNSEKKNVSSITFOKIGA 620
: : : : :
QY 634 AE---KAISLSTEVEN--RFWQTPSSQAPAHAYLSPRTVLVSFFREELGYQARRGGFVG 689
: : : : :
Db 621 FEDELKAV-LPKEVASAALESQNPALPNRITECRSYPLIRFVKEIGT-----ELTIG 674
: : : : :
QY 690 VQOETIGSNVSRITFALIDGRINHVLYKML 719
: : : : :
Db 675 EKVRSPGEBCDKVFTAMCNGQIIDPMLECL 704
: : : : :

RESULT 8

JQ2265

phenylalanine ammonia-lyase (EC 4.3.1.5) - western balsam poplar x cottonwood

C/Species: Populus trichocarpa x Populus deltoides (western balsam poplar x cottonwood)

C/Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999

C/Accession: JQ2265

R/Substratium, R.; Reinold, S.; Molitor, E.K.; Douglas, C.J.

Plant Physiol. 102, 71-83, 1993

A/Title: Structure, inheritance, and expression of hybrid poplar (Populus trichocarpa x

A/Reference number: JQ2265; WUID:94151434; PMID:8108506

A/Accession: JQ2265

A/Molecule type: mRNA

A/Residues: 1-715 <SUB>

A/Cross-references: GB:111747; NID:g169453; PIDN:AAA33805.1; PID:g169454

A/Note: the authors translated the codon AAG for residue 331 as Glu and AAG for residue

C/Comment: The enzyme plays a role in linking primary metabolism to phenylpropanoid meta

C/Superfamily: histidine ammonia-lyase

C/Keywords: ammonia-lyase; carbon-nitrogen lyase

F/201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted

F/202/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.1%; Score 1016; DB 2; Length 715;
Best Local Similarity 35.8%; Pred. No. 1.8e-59;

Matches 266; Conservative 132; Mismatches 276; Indels 68; Gaps 18;

QY 8 LATTLANGFTNGSHAAFTKSAAGPTSAALRTPGLDGHAHQSOLEIVQELSDPTDDVYE 67
: : : : :
Db 1 METVTNGQNGSLSLSCVNGQDPLSM-----GVAAEAMKQSHLDEVKRWADRRKPVYX 55

QY 68 LSGYSTVADVVG--AARKGRRVYQNDDEIRAVDKSVDFLKAQIONSV--YGVTTGFGG 124
: : : : :
Db 56 LGGETITLIAQVASIGHDGTGVKVELSEARPGVAVASDWMDSMDKGTDSYGVTTGGA 115

QY 125 SADPTREDVSLQKALIEHQLOGVTPTSXSFVGRGLE--NTLPLEVVGAMVIRVNSL 182
: : : : :
Db 116 TSHRTKKGALQKELIRFLNAGT-----FNGTETCHTLPHSATRAAMLVRINTL 166

QY 183 TRGSAVRLVLEALTNFNLHRTITPVLKGSISASGDLSPSYIAGATGHPDKVYVHL 242
: : : : :
Db 167 LQYSGIRFEILEAITRLNNITPCLPLRGTTASGDVPLSYIAGLTGRPNKSA--- 223

QY 243 HEGTEKIMFARBAISLFGLEA--VVLGKREGGLVNGTAVSASMATLSLHDSHMLSLSQ 300
: : : : :
Db 224 TGPTEGVLDAAAFKAAIGESGFPELQKEGALVNGTAVSGLASVWLFETNVLAVSE 283

QY 301 ALTLAVEAMVGGQGSFAPFIDVCRPHQGVAVANIRITLSSGSSFAVHEBEVK 356
: : : : :
Db 284 LLSAIFAEVNNKGP-EFTDHLTHKLKHHPQIEAAIMEHILDGSAVWKAKKHETDPL 342

QY 357 VKDDEGLRQDRYPLRTSPQFLGPEVDMMAVSTLSLNNTTDNDPLDVEKQOTAGH 416
: : : : :
Db 343 QKP-----KQDRYALRTSPQFLGQIEVIRATKMIREINSVNDNPLIDVSNKKAHGN 397

QY 417 NPQASAVSISMEKTRIALALIGKINFTQCTELLNAAAMRGIPSL-AAEDPSINHGKGL 475
: : : : :
Db 398 NFGPTPIGVMNDVRLALASIGKIMFAQFSELVNDYNNGLPSNLTASRNPISIDYFG 457

QY 476 DIHIAAYASELGHANPTTVTPVQEMKNOAVNSIALISARTAEANDVLSLLASHLYC 535
: : : : :
Db 458 EIMASVYSELQYLANPTTVTHVQSAEHNQDVNSLGLISSRRTAEAVDILKLMSSTYFLVA 517

QY 536 TLQAVDLRAMELDFKKQFDELLPTLLQOHLGTGLDVNALALEYKALNKRLEQ----- 588
 Db 518 LQAIDLRHLEENLKSAVKNTASQVSRVLTGTGANGE---LHPSRFCEKELIKVDRREV 574
 QY 589 -----TTTYDEPFHMDHAFSYATGTVEILLSSPSANVLTAVNMAKVASAKAIS- 639
 Db 575 FAYVDDPCSATYPLMQKLRQVF-----VDHALENGNEKNFSTSV-FQKIAEFEEELKA 627
 QY 640 -LTREVRN-RFWQTPSSQAPAHAYLSFRTVELYSFVREELGVQARAGDVFGVQOETIGS 697
 Db 628 LLEPEVSARAAVDSSGSAIDNKKICECRSPYLKFFREELGV-----LTGEXKVSDFGE 682
 QY 698 NVSRITRYEALDGRINHYLVKML 719
 Db 683 EFDKVFAMQCGKIIDPMECL 704

RESULT 9

phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
 N/Aleternate names: protein T4D2.190
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
 C/Accession: T46172; S52991
 R/Nyakatura, G.; Fairmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
 submitted to the Protein Sequence Database, December 1999
 A/Reference number: Z23025
 A/Accession: T46172
 A/Molecule type: DNA
 A/Residues: 1-717 <ENYA>
 A/Cross-references: EMBL:AL132958; NID:6434245; PIDN:CA664229.1; PID:g6630746
 R/Experimental source: cultivar Columbia; BAC clone T4D2
 R/Name: L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
 Plant Mol. Biol. 27, 327-338, 1995
 A/Title: The phenylalanine-lyase gene family in Arabidopsis thaliana.
 A/Reference number: S52990; MUID:95195160; PMID:7886622
 A/Accession: S52991
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-205; 'H', 207-234, 'E', 236-717 <MAN>
 A/Cross-references: EMBL:U33678; NID:9497420; PIDN:AAC18871.1; PID:g497421
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C/Genetics:
 A/Gene: PAL2
 A/Map position: 3
 A/Intons: 135/2
 A/Note: T4D2.190
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/203-205/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/204/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.1%; Score 1015.5; DB 2; Length 717;
 Best Local Similarity 36.0%; Pred. No. 1.9e-59;
 Matches 264; Conservative 140; Mismatches 284; Indels 45; Gaps 17;
 QY 5 LDSIATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLGHAHQSGLEIYOEILLSPDTPDD 64
 Db 1 MDQIEAMLGGGSEKTKVAATTKTLADPLNM-----GLADQMKSHLDEVKKMVEBRRP 55
 QY 65 VVELSGYSLTRDVVGAARKGRVRVQNDERIRAVDKSVDFLKAQLQNSV--YGVTTGF 122
 Db 56 VVMIGETILLICGVAAISITVGSVKEVLAETSRQVKAASDVMMSNMKGTDSYGVTGTF 115
 QY 123 GGSADRTEDAVSLQKALLHQLGCVTPTSXSSPFGVGLENTLPLEVVRGAMVIRNSL 182
 Db 116 GATSHRRTKNGTKLQCELIRFLNAGT-----FNGTETCTHTPQSATRAAMVVRVNTL 168
 QY 183 TRGSAVRLVLEALTNPLNHRITPVLPRGSISASGDLSPSLYAGAITGHPVVKVHL 242
 Db 169 LQGYSGIRFEILIAITSLNHNISPLPLRGITITASGDLVPLSYAGLITGRPSKA--- 225
 QY 243 HEGTE-KIMFARBAISLFGLEA--VTLGPKEGELGLVNGTAVSASMAITSLHDSIMLSLS 299

Db 226 -TSPDESILTAAEAEKAGISTGFPDLPQKEGLALVNGTAVGSAWVLEFNAVQAVLA 284
 QY 300 QALTALTVEAMVQGGGSPAPFIHDYCRPHPGQVFAARIRTLTSSGSPAVHEEYVKD 359
 Db 285 EYLSALFAEVMNGKP-EFTDHLTRLKHPPQIEAALMEHILDGSSY-MQLAQVHEMD 342
 QY 360 DEGIIRQDREYPLRTSPQFGLPVEDMHAYSTLSLENTTTDNDPLDVENKOTAHGNFQ 419
 Db 343 PLQKRPQDRAVLRTPSPQWLGPEIVIRQATSIEREINSVNDNLLIVSRKAIHGNFQ 402
 QY 420 ASAVSISMEKTRPLALILGKLNFTCTLLAANRGIPCL-AAEDPSLNYHKGIDIH 478
 Db 403 GTPIGVSMNTRLAALAIKLMFAQFSELVDFYNNGLPSNLTASSNPSLIDYGFKAELA 462
 QY 479 IAAVASELGHANPVTTFVQAPEMNGQAVNSLALISARTAEANDVLSLASHLYCTLLQ 538
 Db 463 MASYCELOYLANPVTSHVQSEHQNOVNSLGLISSKTSIBAVDILKMSSTTLVGICQ 522
 QY 539 AYDLRAMELDFKKQFDELLPTLLQOHLGTGLDVNALALEYKALNKRLE-----EQTTY 592
 Db 523 AYDLRHLLEENLQTVKNTVSGVAKKVLTTGNGE---LHPSRFCEKDLKVDREQVFTY 579
 QY 593 DLEP--RMDASVATGTVVE-LSSSPSANVTLTAVNMAKVASAKAIS--LTREVRN 647
 Db 580 VDDPCSATYPLMQKLRQVTVDAUSNGETERNAVTSI-FQKIGAFEEELKAVLPKEVEAA 638
 QY 648 FMTDPSQAPAHAYLSP-RITVELYSFVREELGVQARAGDVFGVQOETIGSNVRIEAT 706
 Db 639 RAAYNGCATPIRNRKICEGRSPYLRFVREELGTLLIGEXV-----SPGEFPDXVTAM 693
 QY 707 XQGRINHYLVKML 719
 Db 694 CEKGLIDPLMDCL 706

RESULT 10

S52991
 phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - soybean
 C/Species: Glycine max (soybean)
 C/Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C/Accession: S22991
 R/Frank, R.L.; Vodkin, L.O.
 DNA Seq. 1, 335-346, 1991
 A/Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max
 A/Reference number: S22991; MUID:92190550; PMID:1799682
 A/Accession: S22991
 A/Molecule type: DNA
 A/Residues: 1-713 <FRA>
 A/Cross-references: EMBL:X52953; NID:618376; PIDN:CA97129.1; PID:g18377
 C/Genetics:
 A/Gene: PAL1
 A/Intons: 131/2
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/199-201/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/200/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.0%; Score 1011; DB 2; Length 713;
 Best Local Similarity 36.9%; Pred. No. 3.8e-59;
 Matches 270; Conservative 120; Mismatches 282; Indels 60; Gaps 19;

QY 14 NGFTGSHAAPTKSAAGPTSAALRRTPGIDGAAHQSGLEIYOEILLSPDTPDVELSGYSL 73
 Db 5 NCHQNGSFCIST-AKGNNDPLNM--GAALAMKSHLDEVKKMVAEYKRPVRLGGFTL 60
 QY 74 TVRDVGAARKGRVRVQNDERIRAVDKSVDFLKAQLQNSV--YGVTTGFGSADRTTE 131
 Db 61 TIAQVAAYAGHDHGAVALSESAREGVKASSEWMSNMNGTDSYGVTGTFGATSHRRTK 120
 QY 132 DAVSLQKALLHQLGCVTPTSXSSPFGVGLENTLPLEVVRGAMVIRNSLITGHSVRL 191
 Db 121 QGGALQKELIRFLNAGT-----FNGTSSSHLTPITRAAMVVRINLTLLQGYGIR 173

Query March 27.9%; Score 1006; DB 2; Length 716;
Best Local Similarity 36.0%; Pred. No. 8.2e-59;
Matches 266; Conservative 130; Mismatches 275; Indels 68; Gaps 18;

RESULT 13

RESULT 14

A; Experimental source: cultivar Samsun NN; tissue-type leaves after infect

C:Genetics:
 A:Gene:PAL
 C:Function:
 A:Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
 A:Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicitor
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 C:KeyWords: ammonia-lyase; carbon-nitrogen lyase
 F:198-200/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:199/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.8%; Score 1005; DB 2; Length 712;

Best Local Similarity 37.4%; Pred. No. 9.5e-59;

Matches 260; Conservative 131; Mismatches 253; Indels 52; Gaps 17;

QY 49 SOLEIYQELISPTDVEVLSGLTVDPVVGARK--GRRVQNDDEIRARVDKSYDF 106
 DB 33 SHLDEKVAERKRVKVLGGETITVACVAAIAKDKAKVKEISGAAAGYKASDW 92
 QY 107 LKAQLQNSV--YGVTTGFGSADRTEDAVLSQKALIEHQLOGVTPTSXSSFSVGRL 163
 DB 93 VMDSMKGTDSYGVTTGFGATSHRRTKNGALQKELIFLNAGV-----FNGTRES 143
 QY 164 -NTLPLEVRGAMVIRVNSLTGRGSAVLVLEALTNLNRIPTIPIVLRGISASGDL 222
 DB 144 CHTLPGSGTRAMLVRIINTLLQYSGIRFELIEATIKLNHVPLPRLRGTITASGLV 203
 QY 223 PLGYIAGATGHPDVVAVLHEGTEKIMFARPAISLFGLEA--VVLGPKGELGVNGTAV 280
 DB 204 PLSTIAGLITGRPNKA--IGPNG--ETLNAEAFVAGVNSGFEELQKEDALVNGTAV 260
 QY 281 SASMATLSLHDSMTLSLSQALTALTVEAMVQGGSPAFPHDVCRRHPGQEVANIR 340
 DB 261 GSGIAEMVIFDANITLAVFSEVSAIFAEVWNGKP-EFTDHLTKLKHHPQOIEAAIMEH 319
 QY 341 LLSGSEFAVEHEEVEVKVCDGEGIRQDRPLRTSPQFGLVEDMGMHAYSTLSLNNTT 400
 DB 320 ILDGSSY-VKAPQKHETPLQPKQDYALTSQMLGPOLEVRKATKIEREINSVN 378
 QY 401 DNEILDVENKQTAAGNFPQASAVSISMEXTRLALIGKLNFTQCTELNNAANGLPSC 460
 DB 379 DNEILDVSNKALHGNFQGTPIGVSMVARILALASIGKLMFAGSELVNDYNNGLPSN 438
 QY 461 L-AAEDPSLTHGKGLDITIAAYASGLHLANPYTTPQAPAMGQAVNSLALISART 519
 DB 439 LTRGRNPSLDYGRKSEIMASVCELOPLANPVNHHVQSAEQHNQDVSIGLISART 498
 QY 520 EANDVLSLASHLYCTLOAVDRAVELDFKQFDPFLPTLLOCHLGTGLVNALALEVK 579
 DB 499 EAVDILKMSSTLVLCQALIDLRHLEENLRANAVENTVSQVAKRTLTMGANGE--LHPS 555
 QY 580 KALNKRLQCTTYDLEPRHDAFSYATGVELL-----SSPSANVTTLAVN 627
 DB 556 RFCEKDLIRVVDREYFRYADACSNAPYLMQKLOVLDHALENGEMKANSIFQ-- 613
 QY 628 AMVVASAE--KAISTREVRN-RFMQTSQAPAHAYLSPRTVLVSFVBEELGVQARR 683
 DB 614 --KILAFEGELKAV-LPKEVESARISLENGNPALARIKECRSYFLYFVREELG----- 665
 QY 684 GDVAVQOCTTIGSNVRIYEAKDGRINHVLVKML 719
 DB 666 AELLTGEKVRSPGEBQKVFTAMCNGQIIDSLECL 701

RESULT 15

S48725

phenylalanine ammonia-lyase (EC 4.3.1.5) 2 - parsley
 C:Species: Petroselinum crispum (parsley)

C:Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999

C:Accession: S48725

R:Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
 Eur. J. Biochem. 225, 491-499, 1994

A:Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase iso
 A:Reference number: S48725; PMID:95010141; PMID:7925471

A:Accession: S48725
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-716 <APP>
 A:Cross-references: EMBL:X81158; NID:9534892; PIDD:CA547056.1; PIR:9534893
 C:Genetics:
 A:Gene: PAL2
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 F:202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:203/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.8%; Score 1005; DB 2; Length 716;

Best Local Similarity 35.8%; Pred. No. 9.6e-59;

Matches 264; Conservative 133; Mismatches 271; Indels 70; Gaps 19;

QY 12 LANCFTNSHAPTSAGPTSAALRTP---GLDGHAAHQSOLEIYQELISPTDVEV 67
 DB 8 ITNGVANGN-----GMDFCMTKEDPLYWGIAAEAMGSHLDEKVAERKRVK 58
 QY 68 LSGSLTVRADVVG--AARKRRVRVQNDDEIRARVDKSYDFLKAQLQNSV--YGVTTGFGG 124
 DB 59 LGGELLTISQVVAISARQSGVTVLSEAAARAGYKASDWMDSMKGTDSYGVTTGFGA 116
 QY 125 SADRTEDAVSLQKALIEHQLOGVTPTSXSSFSVGRLNTPLEVRGAMVIRVNSLTR 184
 DB 119 TSHRRTKQAGALQKELIFLNAGI-----FNGSDNTLPPSARPAALVRIINTLLQ 169
 QY 185 GSAVALVLEALTNLNRIPTIPIVLRGISASGDLSPSTIYIAGATGHPDVVAVLHE 244
 DB 170 GYSGIRFELIEAITFNLQNTITPCLIRGTTASGDLVPLSTIAGLITGRPNKA--VG 226
 QY 245 GTEKIMFARPAISLFGLEA--VVLGPKGELGVNGTAVSASMATLSLHDSMTLSLSQAL 302
 DB 227 PTGVILSPFAEAKLAGVGGFEELQPKKGLALVNGTAVAGSMAEMVIFEAAILVLAIEVM 286
 QY 303 TALYTEAMVQGGSPAFPHDVCRRHPGQEVANIRLTLSSGSEFAVEHEEVEVKVCDG 362
 DB 287 SAIFAEVWNGKP-EFTDHLTKLKHHPQOIEAAIMEHILDSAV-VKAAQKHEMDPLQ 344
 QY 363 IIRQDRYPLRTSPQFGLVEDMGMHAYSTLSLNNTTDNEILDVENKQTAAGNFPQASA 422
 DB 345 KPKQDRYALRTSPQMLGQILEVIRSTKIEREINSVNDNEILDVSNKALHGNFQGT 404
 QY 423 VGISNEKTRIALALIGKLNFTQCTELNNAANGLPSCLA--AEDPSLTHGKGLDITIA 481
 DB 405 IGSMNTRILATAIAIGKLMFAQFSELVNDFYNNGLPSNLSCGRNPSLDYGRKGAIAVAS 464
 QY 482 YASELGHLANPYTTPQAPAMGQAVNSLALISARTAEANDVLSLASHLYCTLOAVD 541
 DB 465 YGSELOFLANPVNHHVQSAEQHNQDVSIGLISRKTEAIVEILKMSSTLVLCQALID 524
 QY 542 LRAVELDFKQFDPFLPTLLOCHLGTGLVNALALEVKKALNKRLQ----- 588
 DB 525 LRLEENLKSATKYNVSVAKRVLTMG--VNG-ELHPSRFCEKDLIRVVDREYIFAYID 581
 QY 589 --TTTYDLEPRHDAFSYATGVVE--LLSSPSANVTTLAVNMAKVASAKAIS--LTR 642
 DB 582 PGATYPLMQKLRQ-----TLVEHALKNQDNERNMTSIFQ-KIATFEDELKALIPK 632
 QY 643 EYVN-RFMQTSQAPAHAYLSPRTVLVSFVBEELGVQARRGDVAVQOCTTIGSNVR 701
 DB 633 EYVSAAALAESGNPALPKNIECRSYPLKFKRLGT-----EVLTEKXTSPGEBDX 687
 QY 702 IYEAIDGRINHVLVKML 719
 DB 688 VFLAMSGEIIIDLECL 705

Search completed: September 9, 2004, 10:08:45
 Job time : 32.8797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:50:23; Search time 16.4315 Seconds
(without alignments/sec
2281.620 Million cell updates/sec

Title: US-09-939-408a-13
Perfect score: 3610
Sequence: 1 MAPSDSLATLANGFTNGS.....RIYAIKGRINHYVAKLA 720

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261.4	72.4	716	1	PALY_RHOTO
2	2499.5	69.2	713	1	PALY_RHOTO
3	1346.5	37.3	740	1	PALY_AMANT
4	1247.5	34.6	724	1	PALY_USITMA
5	1035.5	28.7	725	1	PALY_ARATH
6	1024	28.4	717	1	PALY_PRUV
7	1020	28.3	707	1	PAL4_ARATH
8	1019	28.2	725	1	PALY_TRIST
9	1017.5	28.2	718	1	PAL3_PPCOR
10	1016.5	28.2	715	1	PAL1_TOBAC
11	1016	28.1	715	1	PALY_POPTR
12	1015.5	28.1	717	1	PAL2_ARATH
13	1013.5	28.1	713	1	PALY_DIGLA
14	1012	28.0	714	1	PALY_CAMST
15	1011	28.0	713	1	PAL1_SOYAN
16	1009	28.0	725	1	PALY_MEDSA
17	1006	27.9	716	1	PAL1_PETCR
18	1005.5	27.9	723	1	PAL1_PEA
19	1005	27.8	712	1	PAL3_TOBAC
20	1005	27.8	716	1	PAL2_PETCR
21	1004	27.8	715	1	PALY_STYHU
22	1003.5	27.8	712	1	PAL2_PHAVU
23	998	27.6	701	1	PAL1_ORYSA
24	998	27.6	705	1	PAL2_LITTR
25	997.5	27.6	721	1	PAL5_LYCES
26	997	27.6	720	1	PAL1_SOLTU
27	995.5	27.6	724	1	PAL1_LYCES
28	995.5	27.6	724	1	PAL2_PEA
29	995	27.6	710	1	PAL1_LITTR
30	993.5	27.5	708	1	PAL2_TPOBA
31	993	27.5	712	1	PAL2_TOBAC
32	991	27.5	718	1	PAL2_CICAR
33	987.5	27.4	667	1	PALY_HELAN

34	983	27.2	708	1	PAL1_DAVCA	O23865 daucus caro
35	982.5	27.2	710	1	PAL2_ORYSA	P53443 oryza sativ
36	980	27.1	710	1	PAL1_RUBID	O98568 rubus idaeu
37	979.5	27.1	722	1	PALY_CITUL	Q42667 citrus limo
38	973.5	27.0	700	1	PALY_WHEAT	Q43210 triticum ae
39	973.5	27.0	754	1	PALY_PINTA	P52777 pinus taeda
40	966	26.8	590	1	PAL2_SOLTU	P31426 solanum tub
41	964.5	26.7	695	1	PAL3_ARATH	P45725 arabidopsis
42	954.5	26.4	707	1	PAL1_TPOBA	P14166 ipomoea bat
43	953.5	26.4	710	1	PAL2_POPKI	Q43052 populus kit
44	947	26.2	682	1	PAL1_POPKI	P45731 populus kit
45	946.5	26.2	710	1	PAL3_PHAVU	P19143 phaseolus v

ALIGNMENTS

```

RESULT 1
ID PALY_RHOTO STANDARD; PRT; 716 AA.
AC P15344;
DT 01-OCT-1989 (rel. 12, Created)
DT 01-NOV-1990 (rel. 16, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
PAL.
OS Rhodospiridium toruloides (Yeast) (Rhodotorula gracilis).
OC Eukaryota; Fungi; Basidiomycota; Uredinomycetes;
OC Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
CX NCBI_TaxId=5286;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 14;
RX MEDLINE=92127061; PubMed=1773059;
RA Rasmussen O.F., Orum H.;
RT "Analysis of the gene for phenylalanine ammonia-lyase from
  Rhodospiridium toruloides."
RL DNA Seq. 1:207-211 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 0559;
RX MEDLINE=88112870; PubMed=2828184;
RA Anson J.G., Gilbert H.J., Oram J.D., Minton N.P.;
RT "Complete nucleotide sequence of the Rhodospiridium toruloides gene
  coding for phenylalanine ammonia-lyase."
RL Gene 58:189-199 (1987).
CC -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
  phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -!- PATHWAY: Phenylpropanoid biosynthesis, first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
  which is formed autocatalytically by cyclization and dehydration
  of residue Ala-Ser-Gly (By similarity).
CC -!- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC
DR EMBL; X51513; CAA5886.1; -
DR EMBL; M18261; AAA33883.1; -
DR EMBL; X12702; CAA31209.1; -
DR PIR; A29607; A29607.
DR PIR; A56628; A56628.
DR HSSP; P21310; 1B87.
DR InterPro; IPR008948; L-Asparase-like.
DR InterPro; IPR001106; Phe_Am_NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.

```

DR Pfam, PF00221; PAL, 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism.
FT CROSSLINK 211 213 5-Imidazolinone (Ala-Gly)
FT MOD_RES 212 212 DHA (2,3-DIHYDROXALANINE)
FT CONFLICT 4 37 (BY SIMILARITY).
FT RPTSQSQARTC (IN REF. 2).
SQ SEQUENCE 716 AA; 76879 MW; 0CIDF6176944E5E6 CRC64;

Query Match 72.4%; Score 2614; DB 1; Length 716;
Best Local Similarity 72.0%; Pred. No. 2,4e-167;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
1 MAPSLDLSATTLANGFTNGSHAAPTSAAGPTSLARTPTGLDGHAAHOSOLEIVQELISD 60
1 MAPSLDLSHSPANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTDVIVKMLAA 54
61 PTDDVVELSGYSLTRVDVGAARKRRVQNDDEIRAVDKSVDFLKAQLONSYGYTT 120
55 PTSTLELDGYSLNGVYSAARKRPVVKOSDEIRSKIDKSVFELRSQLSMSYGYTT 114
121 GFGGSADRTEDAVSLQKALIEHQLCGVPTXSXSPSVRGENTLPLEVVRGAVIRVN 180
115 GFGGSADRTEDAVSLQKALIEHQLCGVPTXSXSPSVRGENTLPLEVVRGAVIRVN 174
181 SLTRGSHAVLVLEALTNFLNRITPIVPLRSGISASGDSLPLYAGAITGHPDVXV 240
175 SLTRGSHAVLVLEALTNFLNRITPIVPLRSGISASGDSLPLYAGAITGHPDVXV 234
241 VLHGTGKIMFAREALSLRGLEAVLGRKGGVGVNGTVASMTLSLHSHMLSLISQ 300
235 VVEHGKELLYAREAMALFNLBPVVGPRKGLVNGVASMVTLALHAAHMLSLISQ 294
301 ALTALTEAVMGOQGFAPFIDVCPHFGQVEVARNRTLLSGSPFAVEHEEVKVD 360
295 SLTMTVEAVMVGAGSHFPLHVTMRPHPTQLEVAGNIRKLEGGSPFAVHEEVKVD 354
361 EGIIRQDRYPLRTSPQFLGFLVEDMHAAYSTLSLNTTNDPLLDVENCQTAKGNGFQ 419
355 EGIIRQDRYPLRTSPQFLGFLVEDMHAAYSTLSLNTTNDPLLDVENCQTAKGNGFQ 414
420 ASAVSISMEKTRIALALIGKLTNCTEELNANMGRGLPSCLAADPSLNYHGKLDHI 479
415 AAAYANTMEKTRIGLQIGKLTNCTEELNANMGRGLPSCLAADPSLNYHGKLDIAA 474
480 AAAYASELGLANFVTTFVQPAENGAVNSLALISARTAEANDVLSLLASHLYCTLQA 539
475 AAAYSELGLANFVTTFVQPAENGAVNSLALISARTAEANDVLSLLASHLYCTLQA 534
540 VDIKAMELDFKQEDLPTLLQOHG---TGADV-NLALAEVKALKRLEQTTVYLE 595
535 IDIRALFEFKKQFGPIVSLIDQHGSAFTSNRLDELVEVNTTLLKRLQNTSYDLV 594
596 PRMHDAFSYATGVELLSLSSPSANVTLTAVANAKVASAKAKASLSTREVRNRFQPSQ 655
595 PRMHDAFSYATGVELLSLSSPSANVTLTAVANAKVASAKAKASLSTREVRNRFQPSQ 651
656 APAAAYISPTRTVLVSFVARELQVQARRDVFVQOQTIGSNVRIRYEAIKQGRINVL 715
652 SPALSYLSPRTQILYAFVEELGVKARRDVFELGQEVLTIGSNVSKYIEAIKGRINVL 711
QY 716 VKOLA 720
DB 712 LKOLA 716

RESULT 2
PALLY RHORB
ID PALLY RHORB
AC P10248;
STANDARD; PRT; 713 AA.

DT 01-VAR-1989 (Rel. 10, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL.
OS Rhodotorula rubra (Yeast) (Rhodotorula mucilaginosa).
OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
OC Rhodotorula.
OX NCBI_TaxId=5537;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-15597;
RX MEDLINE=89083582; PubMed=3205749;
RA Filipula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT "Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula rubra."
RL Nucleic Acids Res. 16:11381-11381(1988).
RN [2]
RP DETAILS.
RC STRAIN=NRRL Y-15597;
RX MEDLINE=89083583; PubMed=3205750;
RA Filipula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT "DNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal genomic intron deletions."
RL Nucleic Acids Res. 16:11382-11382(1988).
CC - FUNCTION: Catalyzes the nonoxidative deamination of L-phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO), which is formed autocatalytically by cyclization and dehydration of residues Ala-Ser-Gly (by similarity).
CC - SIMILARITY: Belongs to the PAL / histidine family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X13094; CAA31486.1; -.
CC PIR; S01999; S01999.
DR HSSP; P21310; 188F.
DR InterPro; IPR008948; L-Asparagine-lyase.
DR InterPro; IPR001106; Phe/His NH3-lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism.
FT CROSSLINK 217 219 5-Imidazolinone (Ala-Gly)
FT MOD_RES 218 218 DHA (2,3-DIHYDROXALANINE)
FT CONFLICT 4 37 (BY SIMILARITY).
SQ SEQUENCE 713 AA; 76000 MW; 6EB8317CEB037D88 CRC64;
Query Match 69.2%; Score 2499.5; DB 1; Length 713;
Best Local Similarity 69.6%; Pred. No. 1.1e-159;
Matches 511; Conservative 83; Mismatches 105; Indels 35; Gaps 9;
1 MAPSLDLSATTLANGFTNGSHAAPT-----KSAAGPTSLARTPTGLDGHAAHOSOLEI 53
1 MAPSLDLSATSVANSLSGHAAAAANGDVHKHTAGAGSLPTT-----ETQGLDI 52
QY 54 VOELLSDP-TDDVVELSGYSLTRVDVGAARKRRVQNDDEIRAVDKSVDFLKAQLQ 112
DB 53 VERIADAGANDQILDGYTLTLDVGAARKRSVYKADSPHIREIDISVEFLRTQLD 112
QY 113 NSGVGTTGFGGSADRTEDAVSLQKALIEHQLCGVPTXSXSPSVRGENTLPLEVVR 172

```

Dh 113 NSVYGVTTFGSGSADRTREDAISLQKALLEHQLCGVLTSMWDGFLSGLENSLPLEVVR 172
Qy 173 GAVVTVNSLTGSHGAVNLVLEALTNLNRITPIVLRGSGISAGSLSTAGAT 232
Dh 173 GAVTIVNSLTGSHGAVNLVLEALTNLNRITPIVLRGSGISAGSLSTAGAT 232
Qy 233 GHEDVAVVHTEGTEKIMFARBAISLFGLEAVVLPKXGLGVNGTAVASAMATLSLSDS 292
Dh 233 GHEDSKVHV--DG--KIMAGPAAIKLQIPVYLPKXGLGVNGTAVASAMATLSLSDS 292
Qy 293 HMLSLISQALITLVAVAMVGOQGSFAPFIHDVCRPHQGVAVARNITRLSSGSPAVEHE 352
Dh 289 HVLSLAQALITLVAVAMVGHAGSFHPFIHDVCRPHQGVAVARNITRLSSGSPAVEHE 348
Qy 353 EEKVVDDSGILLRORRYPRTSPQELGPIVEMMGASTLSLE--NNTTNDPILDVENKQ 411
Dh 349 TEKVVDDSGILLRORRYPRTSPQELGPIVEMMGASTLSLE--NNTTNDPILDVENKQ 408
Qy 412 TAHGNFQASAVSISMETRLALATLIGKLNFTQCTELLNAANRGLPSCLAEDPSLNYH 471
Dh 409 THHGAFMMSVSGNTEKTRIALVAMGKVSFTQLTEMLNAGNRRALPSCLAEDPSLNYH 468
Qy 472 GKGLDHIHAAVSELGHLNAPVTTFVQPAEMGNQAVNSLALISARTAEADVLSLLAS 531
Dh 469 CKGLDIAAAVSELGHLNAPVTHVQPAEMGNQAVNSLALISARTAEADVLSLLAT 528
Qy 532 HUYCTIQAVDYLAMELDFKQFDPILPTLLOQLLTGLDVNALAL-----EVKALKNRL 586
Dh 529 HUYCTIQAVDYLAMELDFKQFDPILPTLLOQLLTGLDVNALAL-----EVKALKNRL 582
Qy 587 EOTTIVLEPRHMDAFSVATGVIVELLSSPSANVTLTAVNAKVAEKAISLTRYENR 646
Dh 583 QQNNSYDLQRKMDTFSVATGAVVETLAGQ--EVSLSLNAKVAEKAISLTRYENR 639
Qy 647 RFPQGTSSQAPAHAVISPTETVLVSFVEBELGVQARQGVFVGVQOETIGNSVITYEAI 706
Dh 640 SEFAAASSSSPALKYISPTETVLVSFVEBELGVQARQGVFVGVQOETIGNSVITYEAI 699
Qy 707 KQGRINHVLMKLA 720
Dh 700 KSGCIAPVLMKMA 713

RESULT 3
ID PALT AMAMU STANDARD; PRT; 740 AA.
AC 033967;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL.
OS Amanita muscaria (Fly agaric).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Amanitaceae; Amanita.
OC NCBI_TaxID=41956;
RN [1]
RP SEQUENCE FROM N.A.
RP Nehls U., Mikolajewski S., Ecker M., Hamp R.:
RP "Sugar and nitrogen-dependent regulation of an Amanita muscaria
RP phenylalanine ammonium lyase gene.";
RT Submitted (JUG-1998) to the EMBL/GenBank/DBJ databases.
RU
CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PPM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ser-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.embnet.org/licenses/embnet.html).
CC or send an email to license@ebi-sib.ch).
CC
DR EMBL; AJ010143; CA09013.1; -.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Asparaginase-like.
DR InterPro; IPR001106; Phe/His NH3 lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRPFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KM Lyase; Phenylpropanoid metabolism.
FT CROSSLINK 224 226 5-imidazolone (Ser-Gly)
FT (By similarity).
FT MOD_RES 225 225 DHA (2,3-DIDEHYDROALANINE)
FT (By similarity).
SQ SEQUENCE 740 AA; 80156 MW; B055CFE3D8B7BCEE CRC64;

Query Match 37.3%; Score 1346.5; DB 1; Length 740;
Best Local Similarity 43.0%; Pred. No. 2e-82; Indels 51; Gaps 15;
Matches 317; Conservative 131; Mismatches 238;

Qy 10 TILANGF--TNGSHAPTKSAAGPTSAALRRTPGLDGHAAHQSO--LEIYOZLLSPDTDVV 66
Dh 25 TTPVNGFKTALSKASRMT--KTSAL-----SOFLEAVRELEGYKNGRAI 68
Qy 67 ELISGSLTVRDVYGARKGRVYVONDEIRAYDKSIVFL--KAOLNSYVYVTTGGGG 124
Dh 69 KYDGOTLSIAAFAAARVAAVELEDESPLVERKRSQSLAANKVSGASVYGSTGGG 128
Qy 125 SADTRFEDVVSQKALIEHOLGCVPTXSFSV--GRLGINT--LPLEVGAMVIRNS 181
Dh 129 SADTRDKMLGFLLOHNGHILPTSTEPDVLPLQDANTSMPEMIGALLIRNS 188
Qy 182 LTRGSHAVLVLEALTNLNRITPIVLRGSGISAGSLSTAGATGHPDVYHV 241
Dh 189 LTRGSHGIRWELIEKRELLANVPIVPLRGSISSGDLPLSLTIGTIIIGNPSIKYV 247
Qy 242 LH----EGTEKIMFARBAISLFGLEAVVLPKXGLGVNGTAVASAMATLSLSDSHMSL 297
Dh 248 -HGPSKSGIRQIGSSKDVIALNINIEPPLSEKPEPLGLNGTAFSAAALNKAIRHVL 306
Qy 298 LSCALITALVAVAMVGOQGSFAPFIHDVCRPHQGVAVARNITRLSSGSPAVEHEEYKV 357
Dh 307 LQVCTAMGTEALIGTBASHAFIHATARPBGQVECAENIMWLLDGSKLAQLEHEVRL 366
Qy 358 KQDEGLRQDRYPLRTSPQFLGPIVEMMGASTLSLENN--TTTNDPILDVENKQRTHG 415
Dh 367 EDDKTYLRQDRYPLRTSPQFLGPIVEMMGASTLSLENN--TTTNDPILDVENKQRTHG 426
Qy 416 GNFQASAVSISMETRLALATLIGKLNFTQCTELLNAANRGLPSCLAEDPSLNYHKGKL 475
Dh 427 GNFQAMAVTNAMETRLALHNGKLLFSQSTELVVPANRGLPESVAALTDLSNYHAAGL 486
Qy 476 DHIHAAVSELGHLNAPVTTFVQPAEMGNQAVNSLALISARTAEADVLSLLASHLYC 535
Dh 487 DATAVVAE----ATPGPTHIQSAEMENQAVNSLALISARTAEADVLSLLASHLYC 542
Qy 536 TIQAVDYLAMELDFKQFDPILPTLLOQLLTGLDVNALAL-----EVKALKNRL 586
Dh 543 LQCALDLRLQREFFPGDLITIRBELRSSFGSLSEQVEKIQUNYLSAFEDHDKTTIM 602
Qy 593 DLEPRHMDAFSVATGVIVELLSS--PSANVTLTAVNAKVAEKAISLTRYENR 647
Dh 603 DNDTRKTTMAATSSSVLQFFTDSGASVPSPSCDLSVSSQSVAVARRSSVLMDDAKE 662
Qy 648 FVQTPESQAPAHAVISPTETVLVSFVEBELGVQARQGVF-----VGVQOETIGNSVIR 701

```

D6	663	YIFGDSGPPTASQYIG-KTRPVYQFLRTTIGVRKKGSGSNVKKFVNGLGVEDVTITGNISR	722
Oy	702	IYEAIKDGRNHNHLYMCK	718
D6	722	IYESIRDKXMSIIVSL	738
<hr/>			
RESULT 4			
ID	PALY_USDMA	STANDARD;	PRT; 724 AA.
AC	Q96V77;		
DT	10-OCT-2003 (Rel. 42, Created)		
DT	10-OCT-2003 (Rel. 42, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Phenylalanine ammonia-lyase (EC 4.3.1.5).		
DN	PA11		
OS	Ustilago maydis (Smut fungus).		
OC	Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;		
OC	Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.		
OX	NCB1_Taxid=5270;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Kim S.-H., Macdonald K., Vitmani D., Wake K., Kronsted J.W.,		
RA	Ellis B.B.;		
RT	"Cloning and disruption of a phenylalanine ammonia-lyase gene from		
RL	Ustilago maydis."		
CC	Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.		
CC	-1 FUNCTION: Catalyzes the nonoxidative deamination of L-		
CC	phenylalanine to form trans-cinnamic acid and a free ammonium ion.		
CC	-1 CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).		
CC	-1 PATHWAY: Phenylpropanoid biosynthesis, first step.		
CC	-1 SUBCELLULAR LOCATION: Cytoplasmic (probable).		
CC	-1 PPM: Contains an active site 4-methylidene-imidazole-5-one (MIO),		
CC	which is formed autocatalytically by cyclization and dehydration		
CC	of residues Ala-Ser-Gly (By similarity).		
CC	-1 SIMILARITY: Belongs to the PAL / histidase family.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF306551; AAL09388.1; .		
DR	InterPro; IPR008948; L-Aspartate-like.		
DR	InterPro; IPR001106; Phe/His_NH3lyase.		
DR	InterPro; IPR005922; Phe_am_lyase.		
DR	Pfam; PF00221; PAL; 1.		
DR	TIGRFAMs; TIGR01226; Phe_am_lyase; 1.		
DR	PROSITE; PS00468; PAL_HISTIDASE; 1.		
KW	Lysase; Phenylpropanoid metabolismm.		
FT	CROSSLINK 205		
FT	MOD_RES 206		
FT	MOD_RES 206		
FT	MOD_RES 206		
SO	SEQUENCE 724 AA; 79277 MW; 02473697034CDIAB CRC64;		
<hr/>			
Oy	Query Match	34.6%; Score 1247.5; DB 1; Length 724;	
Oy	Best Local Similarity	40.8%; Pred. No. 8.1e-76;	
Oy	Matches 302; Conservative 130; Mismatches 243; Indels 65; Gaps 18;		
D6	23	APTKSAGGTSLRRTPGLDGAHAQSQLEIVEGLSDLPD-----DYVELSGLY 71	
D6	2	APADVLVPVEVSTRPG-----LVQSPDYTKLRASSFRRIQYVIDSY 44	
Oy	72	SLTVRDVVGAAKGGRRVAVONDDERARVDKSVDFLKAQO--NSVYGVTTFGGSADR 129	
Oy	45	NLKIQGLVASAYGVHTPRPBAEAKRKRIDSVQSLLAKLDGESIIYGINFGSGADR 104	
Oy	130	TBDNVSLOKALLEHQLCVTPPSKSSPFVGR-----GLENT-----LPLEVIRGMNV 176	

Db	100	TANTRALQGLALLQWQCGVLVP--STPTPEBPSAPALPLPTDTSLSLIMEAWRGALIV	160
Qy	177	IRPNALSTRHSANVLVLEALTN-FLNHRITLTPYLRGSGISASGDLSPLSYAGITHP	235
Db	164	VRLLSIIMRHSQVMEVLDKMQKFLDNNTTTPVVPKRSSISASGDLSPLSYAGLAQQR	223
Qy	236	DVYVYTHHGCTEKI-MFRRBATSLFGEAVLVLPKREKGLGNVGTASASMATLSHDSHM	294
Db	224	GIYCFPTDGRGRVAVTDEACSRNHKITPTQYSPKALGLINGTAFSASVAGLATYEEN	283
Qy	295	LSLISGALTALTVTEAMVGGQGSFAPRTHDYCRPHQGVAVANITLSSGSFA--VEHE	352
Db	284	LSLITQLTAAVAVEALNKGTDSFAPFHEIARPHPQGISAKFIRAHLSGSLAEHLENE	343
Qy	353	EEVKVWDSCGLRQRYPLRSPQGLPELVEDMMHAYSLTSLENNNTTNDPLDY--ENK	410
Db	344	KAVLSBEDNGTLRQRYTLQTSASVWPGLEDEENAKSVEDEINSTDNPMDPYDDG	403
Qy	411	QTAHGNGFOSAVSISMKRTLALALIGKLNFTQCTELINAAAMNRGLPSCLAA--EDPSLN	469
Db	404	RIHGGNFGAAMMTAAVEKIKLACAGAKTFOQMTELVPAMNRGLPANLASTPDLSTN	463
Qy	470	YHCKGLDTHIAAYASELGLHANPTTTPVQAEKNGAANVSLTISARTEANDVLSLL	529
Db	464	FAKKGINTALASVTSSELFENLPVSTHVQAEAVANQAFNSLLISGRPTLDALFELSLIQ	523
Qy	530	ASHLYCTQAVDLRAMELDFEKQDPDLPTLLQCHGTGLD--VNALALEVKALNRL	586
Db	524	AMSVLYLQAGADIRLQYKVAEQPLTILASHSHSGEMMDTQKQELIAQVYLSMSKRL	583
Qy	587	EQCTTYDLEPRMHDVFSYATGTVELLSGSPSANVT--LTAVANMKVASAEKASILTREV	644
Db	584	DETSXDLRDLRYEYTOPASSVLYRPSSELPSCGGADPLRNLYKVRAGVADTEKIRQV	643
Qy	645	RURFMQTPSSQAAPA-VLSPRTVLVSFVREELGYQARGD--VFYG--VQGETGSN	698
Db	644	TIEFLDNPFA--CHASLILKTKRAYEPRKTLVSYPHKGKLNLEFNGEFPQNTTGGY	700
Qy	699	VSRIYEALIKDGRINHVLTVM	718
Db	701	VSVIYASIRDCGLYMKLSEL	720

RESULT 5

PALI_ARATH STANDARD; PRT; 725 AA.

AC P35510; Q9Z0D6;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).

GN PALI AR12637040 OR T108.22.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; OC eurosids II; Brassicales; Brassicaceae; Arabidopses.

OC NCBI_TaxID=3702;

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Landsberg erecta;

RX MEDLINE=95195160; PubMed=7886622;

RA Wanner L.A., Li G., Ware D., Somssich I.E., Davis K.R.;

RT "The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana."

RT Plant Mol. Biol. 27:327-338 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V., Bell C.R., Ketchum K.A., Lee J.J., Rong G., Van Aken S., Umayam L., Moffet K.S., Cronin L.A., Shen W., Pal G., Van Aken S., Umayam L.,

[illegible]

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF036948; AAC78457.1; -
 DR HSSP; P21310; 1B8F.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRfams; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 203 205 5-midazolone (Ala-Gly)
 (By similarity)
 FT MOD_RES 204 204 DHA (2,3-DIDENHYDROALANINE)
 (By similarity)
 FT SEQUENCE 717 AA; 77999 MW; B84DF90A0BF60B3 CRC64;

Query Match 28.4%; Score 1024; DB 1; Length 717;
 Best Local Similarity 36.2%; Pred. No. 7, 4e-61;
 Matches 262; Conservative 133; Mismatches 269; Indels 60; Gaps 17;

14 NGFTNGSHAAPTKSAAGPTSAIRTP--GLDGHAAHSGLEIVGELLSDPTDVELISG 70
 9 NGHNG-----SVLEPELCIKKDPPLNMGVAETLKSHLDEVKMAEVRKPKVXKGG 61
 71 YSLTVRVVGAARKGRVRYQNDDEIRAVDKSVFLKQQLNSV--VGVTTEGSGSDT 128
 62 ESLTISQVAIAITHDSGVKVELSESARAGKASDWDMSKGTDSGVTTGFATSHR 121
 129 RTEDAVSLQKALIEHOLCGVPTXSXSFVGRGLENTLPLEVVGAMVIRVNSLTRGSA 188
 122 RTKGAAALQKELIRFLNAGVFGSTKES-----GHITPHQATRAMLVRINTLLQVSG 174
 189 VRLVYVLEALTNFLNHRITPVPFRGSIASGDLSPSTYAGKITGHPDKVHVEGTEK 248
 175 IREILLEVITKFNANNVTPCLPRGTITAGCLVPLSYAGLWLRPNPSKA-VGPDG--Q 231
 249 IMPAREALISFGLA--VVIQPEKGLVNGTAVASMATLSLHSHMLSLSQLTALT 306
 232 TLSAAKAFEVGINSGFELQPEKGLALNGTAVSGSLASTVLPFTNLALISLSXIF 291
 307 VEANVGQGSFAFPIHDVCRPHPGQVEVARNIRTLISGSSFAVE---HEEEVKVQDEG 362
 292 AEVVGQKP-EFTDLTHKLKHHPGQIEAAIMEHILDGSSYKAKKLEHDDPQKP--- 347
 363 ILNODRPLRTSPQIPGLVEDMMAHYSTLSLNTTNDPLDVENKQTHAGNFOASA 422
 348 --KODRYALRTSPQWGPQIEVIRVSTKSIEREIDSVNDPLIDSRKKAHGGNFQSTP 405
 423 VSI-SMEXTRIALALIKLNFTQCTELNANRGLPSCIA-AEDPSLYNHGKGLDIIIA 481
 406 IGVSMVTRLAIASIGKLMFAPGSELDVFNNGIPSNLSGGRNSLIDYFKGAIEIMAS 465
 482 YASELGLANPVTTFQPAEMGQANSLALSARTAEANDVSLLSLHLYCTIQAVD 541
 466 YCELOFLANPVNHNQSAOHQVDNSIGLISKRTEAAVILTLMSSTFLVALCOAID 525
 542 LRAMELDFFKQFPLPTLLQCHLTGLDVNALALE-VKALNKLKLEO-----T 589
 526 LRHLEMLRTVANKTVSAQAKRTLTGVNGELHPSRFEKDLTKVDREYVFAVIDPCS 585
 590 TTYDLEPRMHDAASYATGYVELLSSPSANTTLTAVNANKYASAEK-AISLTBEVNRNF 648
 586 ATYPLWQKLRQV-----VEHALTNGENKNASTISIFOKIYAFEEELKVLPLPEVNSAR 639
 649 WQPPSSQAPAHAYLSP-RTRVLYSPVREELGVQARGDVPVQOETIGSVVSRYEATK 707
 640 AALDSSAGVPRNITECRSYPLKFRLEELG-----AEYLTEKVRSGEBCDKVYALIC 694

708 DGRI 711
 695 ECKI 698

RESULT 7
 PAL4_ARATH
 ID PAL4_ARATH STANDARD; PRT; 707 AA.
 AC 09S545;
 DT 16-OCT-2001 (rel. 40, Last Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 10-OCT-2003 (rel. 42, Last annotation update)
 DE Probable phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN A73G10340 OR F14P13.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv, Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salancourt M., Lencze K., Rieger M., Ansoorge W., Unselid M.,
 R Fartman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 R Delany M., Boutry M., Grivell L.A., Macho R., Pulgomech P.,
 R De Simone V., Cholisme N., Artiguenave F., Robert C., Broillet P.,
 R Wincker P., Cattoico L., Weissenbach J., Saurin W., Quetier F.,
 R Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Bernes V.,
 R Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
 R Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 R Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 R Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 R Reichelt U., Scharte M., Schoen O., Barques M., Terol J., Clement J.,
 R Navarro P., Collado C., Perez-Perez A., Ottensmeyer B., Duchemin D.,
 R Cooke R., Laude M., Berger-Liaou C., Purrelli B., Masny D.,
 R de Haan M., Maarse A.C., Alcaraz J.-P., Cortez A., Gasuberta E.,
 R Montfort A., Argitrou A., Flores M., Lignori R., Vitale D.,
 R Mannheim G., Haase D., Schott H., Rudd S., Zaccaria P., Mewes H.-W.,
 R Mayer K.F.X., Kahl S., Town C.D., Koo H.L., Talion L.J., Jenkins J.,
 R Rooney T., Rizzo M., Walts A., Utechtack T., Fujii C.Y., Shea T.P.,
 R Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
 R Pal G., Miltcher J., Sellers P., Gill J.E., Feldbaum T.V.,
 R Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 R Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 R Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 R Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 R Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 R Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana".
 RL Nature 408:820-822(2000).
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (by similarity).
 CC - SIMILARITY: Belongs to the PAL / histidase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AC009400; AAF02809.1; -
 DR HSSP; P21310; 1B8F.

QY 301 ALATLVEANVGGGSAFPIHDVCRPHQOYEVANRITLSSGSAVE---HEEVK 356
D 294 VLSAIFAEVWQGR-EFTDHLTRKHKHHPQIEAAIMEHIIHGSAVYADAKKHEMDPL 352
QY 357 VKODEGILRODRYPLRTSPQIFGLVEDMMHAYSTLSENNNTTNPPLDVENKOTAHG 416
D 353 QKP-----KODRYALRTSPQIMLQPLLEVRSPKSIERINSVNDPILIDSRKALHGG 407
QY 417 NFOASAVSISMEXTRLALALIGLNFOTCTELNMAANGDPSCL-AAEDPSLNYHGKL 475
D 408 NFOGTPIGVMNDTRLALASIGLTPAFQSELVNDPYNNGPLPSNLASRNPDLDFGKGS 467
QY 476 DIHAAVASLGLHAPVTTFQPAEMGNQAVNSLALISARRTAANDVLSLLASHYX 535
D 468 EIMAAVSYSELQYLANPVTHVQSAEONHODVNSGLISSRRTKATIEILOMSTFLIA 527
QY 536 TLQAVDIRAMELDFKKQDFPLPTLLQOHLGTGLDYNALALEVKKALKRLEQTTVDLE 595
D 528 LCGAIDLRHLENLK-----NSVKNTVSQVAKKTLTIGVSGELH 566
QY 596 P-----RHMDAFSY-----ATGVVLEL-----SSPSANVTLT 624
D 567 PSRECEKDLKVVDRH-VFSYIDPCSATYPLAQKRLQVLDHALVNGESEKNSITSIF 625
QY 625 AVNMMKVASAKAIS--LTREVEN-RFMOPTSSQAPAHAYLSPTRVLYSPFVREBELGVA 681
D 626 Q-----KATFEELKTLPEKVSARTAYENGNSITANKINCRESYPLKYVREBELGSL 681
QY 682 RRGDFVGVQOETGNSVRSIYALIDGRINHVLVMT 719
D 682 LTGERVI-----SPGECEDLFTAMCOGKIIDLPLKCL 714

RESULT PETR
ID PAL3 PETR STANDARD; PRT; 718 AA.
AC P45729;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 3 (EC 4.3.1.5).
GN PAL3.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC Apium clade; Petroselinum.
CX NCBI_TaxID=4043;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=95010141; Pubmed=7925471;
RA Appert C., Logemann E., Hahnbrock K., Schmid J., Amrhein N.;
RT "Structural and catalytic properties of the four phenylalanine
ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL Eur. J. Biochem. 225:491-499(1994).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X81159; CAA57057.1; -.
DR PIR; S48726; S48726.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Asparagine-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am-lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfam; TIGR01226; phe_am lyase; 1.
DR PROSITE; PS00488; PAL HISTIDASE; 1.
KM Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLINK 204 206 5-imidazole none (Ala-Gly)
FT (By similarity).
FT MOD_RES 205 205 DHA (2,3-DIDEHYDROALANTINE)
FT (By similarity).
SQ SEQUENCE 718 AA; 78165 MW; 17786451FFD6C35 CRC64;
Query Match 28.2%; Score 1017.5; DB 1; Length 718;
Best Local Similarity 36.5%; Pred. No. 2e-60;
Matches 289; Conservative 128; Mismatches 275; Indels 65; Gaps 19;
QY 14 NGFTNGSHAAPTSAAGPTSAIRTP---GLDGHAAHQOLEIVQELSDPTDDVYELS 69
D 5 NGFTNG-HA---NNGDLCCKEKEDPLNMGVAEALTGSHLDEVKMVAEYKRPVYKLE 59
QY 70 GYSLTVRDVYG-AARKGRVRVONDDEIRAVDKSVDFLKAQLQNSV--VGVTTFGCGSA 126
D 60 GETLITISQYAAISADDSGVKVELSEKRAQVYKASSWVWDNMKKGDSIGVTTGCATS 119
QY 127 DRTEDAVSLQKALLIHQLCGVPTSSXSFSGVGLD---NTLPLEVQAMVIRVNSLT 183
D 120 HARTKQGALQKELRFLNAGI-----FGSGAAGANNLTPLHSATRAAMVIRVNTLL 170
QY 184 RGHSAVLVLEALNPNHRTIPIVLRGISASGDLSPSTYAGATIGHDPKVHVL 243
D 171 QYSGIRFEILEATKFLPNHNITPCPLRGITTSGLVPLSTYAGLTLRPUSKA--V 227
QY 244 EGTEKIMEAREAISLFGLEA--VVLGPKEGLVNGTAVASAMATLSLHSHMLSLDQA 301
D 228 GRTGYTLSPBEAFKLAGVGGFFELQPKREGALVNGTAVASGMAVYLPANILAVLE 287
QY 302 LIALTVEANVGGGSAFPIHDVCRPHQOYEVANRITLSSGSAVEHEEVKAKDDE 361
D 288 MGAIFAEVWQGR-EFTDHLTRKHKHHPQIEAAIMEHIIHGSAV-YKAAQKLEMDPL 345
QY 362 GILRODRYPLRTSPQIFGLVEDMMHAYSTLSENNNTTNPPLDVENKOTAHGPNQAS 421
D 346 QKPQODRYALRTSPQIMLQPLLEVRSPKSIERINSVNDPILIDSRKALHGGNFQGS 405
QY 422 AVSISMEXTRLALALIGLNFOTCTELNMAANGDPSCL-AAEDPSLNYHGKLDIHA 480
D 406 PIGVEMDTRLALALIGLTPAFQSELVNDPYNNGPLPSNLASRNPDLDFGGAELTAMA 465
QY 481 AASLGLHAPVTTFQPAEMGNQAVNSLALISARRTAANDVLSLLASHYXCTLQAV 540
D 466 STCSLEQLFANVTTHVQSAEONHODVNSGLISSRRTKATIEILOMSTFLVGLQAI 525
QY 541 DIRAMELDFKKQDFPLPTLLQOHLGTGLDYNALALEVKKALKRLEQ----- 588
D 526 DIRHLENLKSVKNTVSQVAKRVLWNG--VNG-ELHPSFCEKDLRLVVDREYFAYID 582
QY 589 ---TTTYDLERPHMDAFSYATGTVVELSSPSANVTLTAVMMKVASAKAIS--LTRE 643
D 583 DPCSATYPLMQKLR-----TLVEHALNNGKERNLSTIRQKLAFFEDKALKLPKE 635
QY 644 VNN-RFMOPTSSQAPAHAYLSPTRVLYSPFVREBELGVAQARGDFVGVQOETGNSVSR 702
D 636 VETAAALAESGNPAIPNRIKRCRSYPLKYVREBELG-----EVLIGEKTRSPGEEBEKV 690
QY 703 YEALIDGRINHVLVMT 719
D 691 PTAMSGEIIIDLPLECL 707

RESULT 10
 ID PALI TOBAC STANDARD; PRT; 715 AA.
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 NCBI OR PALB.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4097;
 RX NCBI_TaxID=4097;
 RA Fukaesawa-Akeda T.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
 [2]
 SEQUENCE FROM N.A.
 RC STRAIN=cv. Bright yellow; TISSUE=Callus;
 RC Taguchi G., Sharan M., Gonda K., Yanagisawa K., Shimosaka M.,
 RA Hayashida N., Okazaki M.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 first reaction in the biosynthesis from L-phenylalanine of a wide
 variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 which is formed autocatalytically by cyclization and dehydration
 of residues Ala-Ser-Gly (by similarity).
 CC - SIMILARITY: Belongs to the PAL / histidase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by, and for, commercial
 entities requires a license agreement (See <http://www.ips-sib.ch/announce/>
 or send an email to license@ips-sib.ch).
 CC
 CC EMBL: M84466; AAA4122.1; -.
 DR EMBL: AB008200; BAA22948.1; -.
 DR PIR: S66343; S66343.
 DR HSSP: P21310; 1B8F.
 DR InterPro: IPR008948; L-Asparatase-like.
 DR InterPro: IPR001106; Phe/His NH3lyase.
 DR InterPro: IPR005922; Phe/His NH3lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRfam: TIGR01226; Phe am lyase; 1.
 DR PROSITE: PS00468; PAL_HISTIDASE; 1.
 DR Lyase: Phenylpropanoid metabolism; Multigene family.
 FT CROSSINX 201 203
 FT MOD_RES 202 202
 FT DHA (2,3-DIHYDROALANINE)
 FT (by similarity).
 FT (by similarity).
 FT NCBI_TaxID=3694;
 FT (by similarity).
 SQ SEQUENCE 715 AA; 77780 MW; 03CB4B527394C62 CRC64;
 Query Match 28.2%; Score 1016.5; DB 1; Length 715;
 Best Local Similarity 36.0%; Pred. No. 2.4e-60;
 Matches 270; Conservative 130; Mismatches 259; Indels 91; Gaps 19;
 QY 13 ANGFPTN-GSIAATKSAAPTSLARTPGLDGAHQSLGVQLLSDDPDVLELSGY 71
 DB 3 SNCHVNGEPEFELCKRSADPLNWEKAESLRG-----SHLDVKKQVSEFFKPKVWKLGE 57
 QY 72 SLTVRDVVGAA--RKGRRVYQNDDEIFARVDKSVDFPKQJQNSV--YGVTTGFGGSA 126
 DB 58 SLTVAVVAALAVKDRKANGVKVLESEAAAGVAKASDWMGSMNKGTDSTYGVTTGFGATS 117

QY 127 DRTEDAVSLQKALIEHOLCVTPSPKSSSEVSGLENTLPLEVVRAMVIRNLSLRGH 186
 DB 118 HRRITNGALOKELLRFPLNAGV-----FNGTETSHTEPSPATRAMVIRINTLQGY 170
 QY 187 SAVRLVLEALTNFNLHRTPTVPARGSISAGDLSPLSYAGAITGHPDVKVLHEGT 246
 DB 171 SGIRELEILEATIKLINSITPCLPRLGITASGDLVPLSYAGILTRPNSKA-VGNG- 228
 QY 247 EKIMPAFAEALISFGEA--VVLGPEGLGVNGVAVSASMATLSLHDSHMLSLISQLTA 304
 DB 229 -ETLNAEEAFRAVGVNGGFFELQPEGLATLVNGTAVSGMASVLPFNSNLAVVSEVLSA 287
 QY 305 LVEAMVGGQGSFAFIHIDVGRPHGVAVARNRITLISGSFAVEEHEEYKVDDEGIL 364
 DB 288 IFAEVMNGKP-EFTDLTHYKIKHNGQLEAALVEHLDSSY-VKAQQLHEMDPLQKP 345
 QY 365 RODRYPLRTSPQLGPLEVDDMMHAYSTLSLENTTNDPLDVENKOTAFQGNFQASAVS 424
 DB 346 KQDRVALRTSPQWLGPEVIRATKMTREINSVNDNPILDVSRNALRGNFQGPFG 405
 QY 425 ISMEKTRALALIGLNTQCTELLNAAMNGLPSC-LAEDPSLNVHGGLDHIHAYA 483
 DB 406 VEMDARFALASIGLMPAQPSELVNDYNNGLPSNLTASNPBLDYGFGAETAMASYC 465
 QY 484 SELGLANPVTTFVQAPMGQAVNSLALISARTAEANDVLSLLASHYCTLQAVDLR 543
 DB 466 SELQFLAPVTHVQASQHQNQVNSLGLISAKTAEVNDILKMSSTYVALQALDLR 525
 QY 544 AMELDPKQFDPLETLLOQHGTGLDVNALALVYKALKLEQTTTYDLERWEDA-- 601
 DB 526 HLEENIK-----NA-----VKTVSQAARLTLMGANGELHPARF 560
 QY 602 -----PSYANG-----TYVELSSPSAVVTLTAVAMVAVS 633
 DB 561 CEKELLRYDREYLFAYADDPCSCVPLMQKROYLDVHAMNNESEKANNSSIFQKIGA 620
 QY 634 AE--KAISLREVEN-RFQWTPSQAPAHAYLSPRTVLYSFVREELGVGARGDVFG 689
 DB 621 FDEELKAV-LPEVSAARALESQNPALPNRITCRSPYLFRKXELGT-----ELLTG 674
 QY 690 VQGETIGSVSRVIAIKDGRINHVLVKKL 719
 DB 675 EKVRSPEGECDKVFYAMCNGQITDPMTECL 704
 RESULT 11
 ID PALY POPTR STANDARD; PRT; 715 AA.
 AC P45730;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Populus trichocarpa (Western balsam poplar).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Eucoside I; Malpighiales; Salicaceae; Saliceae; Populus.
 NCBI_TaxID=3694;
 RX NCBI_TaxID=3694;
 RA MEDLINE=94151434; PubMed=8108506;
 RA Subramaniam R., Reinold S., Mollitor E.K., Douglas C.J.;
 RT "structure, inheritance, and expression of hybrid poplar (Populus
 trichocarpa x Populus deltoides) phenylalanine ammonia-lyase genes.";
 RL Plant Physiol. 107:71-83(1993).
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 first reaction in the biosynthesis from L-phenylalanine of a wide
 variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC SEQUENCE FROM N.A.
 RC STRAIN=P. trichocarpa x P. deltoides; TISSUE=leaf;
 RC MEDLINE=94151434; PubMed=8108506;
 RA Subramaniam R., Reinold S., Mollitor E.K., Douglas C.J.;
 RT "structure, inheritance, and expression of hybrid poplar (Populus
 trichocarpa x Populus deltoides) phenylalanine ammonia-lyase genes.";
 RL Plant Physiol. 107:71-83(1993).
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 first reaction in the biosynthesis from L-phenylalanine of a wide
 variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- PPM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: L11747; AAA33805.1; -
 CC HSSP: P21310; 1B8F.
 CC InterPro: IPR008948; L-Aspartase-like.
 CC InterPro: IPR001106; Phe/His_NH1lyase.
 CC InterPro: IPR005922; Phe_am_lyase.
 CC Pfam: PF00221; PAL; 1.
 CC TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 CC PROSITE: PS00488; PAL_HISTIDASE; 1.
 CC KEGG: Phenylpropanoid metabolism; Multigene family.
 CC CROSSLINK 201 203 5-Imidazolimine (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 202 202 DHA (2,3-DIHYDROLMANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 715 AA; 77918 MW; 5B9E837A6E8AF2B6 CRC64;
 Query March 28.1%; Score 1016; DB 1; Length 715;
 Best Local Similarity 35.8%; Pred. No. 2.5e-60;
 Matches 266; Conservative 132; Mismatches 276; Indels 68; Gaps 16;
 QY 8 LATTANGFTNGSHAAPTKSAAGPTSAIRRTPLGDLGHAHQSLIEVELSDPTDVAE 67
 DB 1 MEVTYKGYNGSLIESLVNQRDPLSM-----GVAABAMKSGHLEDEVKRVADYRKPVYK 55
 QY 68 LSGSLTVRVYVG-AARKGRVRVYVNDDEIRARVDSVDFLKAQYQNSV--YGTITGGG 124
 DB 56 LGGETTLTAQVASIAHDTGDVAVELSESARCGVASSWWDSDKDTSDYGVATTGGA 115
 QY 125 SADRTTEDAVSLQKALIEHQLCGVPTFSXSSFSVGKLE--NTLPLEVRGAMVIRVNSL 182
 DB 116 TSHRRTQGGALCKELIRFLNAGI-----FNGTETCTTLPHSATRAAMLVIRNTL 166
 QY 183 TRGSAVRVYVLEALNLFNHRITPVPLRGSISASGDISPLSYTAAITGHPDYKVT 242
 DB 167 LQYSGIRPEILRLNNTNNTPLRGTITASGDIVPLSYTAGLLTGRPNVSKA-- 223
 QY 243 HEGTEKIMFARBAISLFGLEA--VVLGPKRGLGVNGTAVASVATLSLHSHMLSLISQ 300
 DB 224 TGTGTEVLDAAEFKAAAGISGFFELQPKGIALVNGTAVSGLSMWLFETNVLAVISE 283
 QY 301 ALTALTEAVAVGQSGFAPFIDHVCPRHPQVEVANRTLLSGSSFAVE---HEEEVK 356
 DB 284 LLSAIFAEVAVNGKRP-EFTDHLTKLKGHPQGLEAAALMEHILIDGSAVKAACKLHEDPL 342
 QY 357 VKDDEGLRQDRYPLTSPQFLGPIVEDMMAVSTLSLENTTTNDLDDVENQQTAAHG 416
 DB 343 QKP-----KODRKALTSPPQWLGPOLEVRFSFKIERINSVNNPLDIVSRNKALHG 397
 QY 417 NFOASAVSISMEKTRTALALIGKLTNCTCELLNANRGLPSCD-AAEDSLVYHGKGL 475
 DB 398 NFGTPIGVSMNVRLAIALSIGKTLPAQSELVNPFYNGGLPSNLTASRNSLIDYGRGA 457
 QY 476 DIIIAVASELGLANPVTTTQOPAEVNGQAVNSIALISARTAEANDVLSLLASHLYC 535
 DB 458 EIMASVSELYLTAPVTTTHVQSAEOHNDVNSLGLISSRRTAEAVDILKMTSTPLVA 517
 QY 536 TLQAVDRAAMELDFKKQDPLPTLLQCHLTGGLVNLALAEVKKALKRLEQ----- 588
 DB 518 LQALIDRLHLEENLKAANVAVSQVSKRYLTITGANGF---LHPSFCCKELLYVDEEYV 574
 QY 589 -----TTTYDLERPHWDAPSYATGTVVELLSSSPSANVTTLTAVNAAKVASAKAIS- 639

DB 575 FAYVDDPCGATPTPLMOKXROVF-----VDHALNGENGEKXNFSTV-FQITEAREEDLKA 627
 QY 640 -LTREYRN-RFQVCTSSQAPAHAVISPTRTVLVSFVREELGVQARQGVGVQOETIGS 697
 DB 628 LIPKEVESAPRAAYDGSNAIDNKKICEKRSYPLXYFVREELGTV-----LITGERVSPGE 682
 QY 698 NVSRYEATKDRGINHVLWKVL 719
 DB 683 EFDKVFYTMCGSKITDPLVECL 704
 RESULT 12
 PAL2_ARATH
 ID PAL2_ARATH STANDARD; PRT; 717 AA.
 AC P45724; O94KC9; O9SCN5;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
 GN PAL2 OR A73G53260 OR T4D2.190.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salambouat M., Lemcke K., Rieger M., Ansoze W., Unsel'd M.,
 RA Faltmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delany M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choisme N., Artiguenave F., Robert C., Broctier P.,
 RA Schaefer M., Cattolico L., Weissenbach J., Sautin W., Quetier F.,
 RA Wurmbach E., Drizonek H., Etile H., Jordan N., Banger S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Reichelt J., Schaefer M., Schoen O., Barges M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Olfenauerder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Liauro C., Purrelle B., Massey D.,
 RA de Han M., Maiese A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Montfort A., Argiriou A., Flores M., Ligouri R., Vitale D.,
 RA Monnaud G., Haase D., Schoof H., Ruid S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kahl S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts R., Utecherback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haase B., Walts R., Wu D., Peterson J., Van Aken S.,
 RA Pail G., Miltscher J., Sellers P., Gill J.E., Feldlyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asanizu E.,
 RA Sasano S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Metanabe A., Yamada M., Yasuda M., Tabata S.,
 RT 'Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RL Nature 408:820-822(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,

Dd		226	-TGPGGSELTAKENFEKAGISTGFDFDQPREGLATLNGTAAVSGMAAMVLPEANQAVLA	284
Qy		300	QALTLAIVAAWGQQSFPAPIDHVCPRPQGVEVARNRITLLSSGSFAVEHEEYKVQ	359
Dd		285	EVLISAFIEFVWVGSR-FETDLTHRLRGHGQLEAAIMEHLIDGSS-MKLAQRKEMD	342
Qy		360	DEGLIROPARYLFKSPQLGPLVEDMMHAASTLSLENNTTDNPFLDVENKOTAHGNFO	419
Dd		343	PLQKKORRYALFRISPWLGRQGVIRQAOKSLERLRINSVNNDPLVDGRNKAINGNQ	402
Qy		420	ASAVSISEKTRALLALIGKINFTOCTELINAAAMRQLPSCL-AAEDPSLYHGKLDIH	478
Dd		403	GPIPIVSANDNTRLIAAIIGXLMFQPSSELVNDVPNNGLPSNLTAASPILDGFPGAELA	462
Qy		479	IAAVASEGHIANVTAEVVOEAENGQAVNSLLISARPTAEANDVLSLLASHLYCTLQ	538
Dd		463	MMSYSSELQYLANPVTSVOSABEHODVNSLGISSRKTSSENVILKIMSTTFVLGICQ	522
Qy		533	AYDRAMELDPKQFDPLLPTLLIQCHLGTGLDVNALALEVKALKRL-----EQTTY	592
Dd		522	AYDLNHLHENLRQYVKNVTSQVAKKVLTTGINGE---LHPSRFCEKDLIKVVDREQVPT	579
Qy		593	DLEP--RWHDFASTAATGYVE-LISSPSANVTLTVANMKVASARKALS-LTREVRNR	647
Dd		580	VDDPSATPYPMQRLOQYVDHASNETSKNAVTSI-PQKGAFEEELKAVLPXEVA	638
Qy		648	FMQTPSOAPAHAYLSP-RTRVLVSFFREEJGVQARGDPFVGOOETIGSNVRIYEA	706
Dd		633	RAAYNGNRPINPNNIKERSYPLVRFREBELGKLUTGEKV-----SPGEEDKVFTEM	693
Qy		707	KGRINHVLVKML	719
Dd			:::::	
Qy		694	CEGKLIDPLMDCL	706
Dd				
 RESULT 13 PALY-DIGLA ID PALY-DIGLA STANDARD; PRT; 713 AA.				
AC	OZ3924;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DS	Phenylalanine ammonia-lyase (EC 4.3.1.5).			
OS	Digitalis lanata (Foxglove).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytes;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;			
OC	Lamiales; Lamiales; Antirrhinaceae; Digitalideae; Digitalis.			
OX	NCB1_TaxID=49450;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Theoinger C.;			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: This is a key enzyme of plant metabolism catalyzing the			
CC	first reaction in the biosynthesis from L-phenylalanine of a wide			
CC	variety of natural products based on the phenylpropane skeleton.			
CC	- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).			
CC	- PATHWAY: Phenylpropanoid biosynthesis; first step.			
CC	- SUBCELLULAR LOCATION: Cytoplasmic (Probable)			
CC	- PTM: Contains an active site 4-methyliden-imidazole-5-one (MIO),			
CC	which is formed autocatalytically by cyclization and dehydration			
CC	of residues Ala-Ser-Gly (by similarity).			
CC	- SIMILARITY: Belongs to the PAL / histidase family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboratio			
CC	n between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercia			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce			
CC	or send an email to license@sb-sib.ch).			
CC				
CC	-----			
CC	EMBL; AJ002221; CA05251.1; --			
CR				

DR HSSP: P21310; 188F.
 DR InterPro: IPR008948; L-Asparagine-lyase.
 DR InterPro: IPR001106; Phe/His_NH3lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRPFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 198 200 5-imidazolinone (Ala-Gly)
 FT MOD_RES 199 199 (By similarity)..
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 713 AA; 77732 MW; 747C64C65BDCC6A CRC64;
 Query Match 28.1%; Score 1013.5; DB 1; Length 713;
 Best Local Similarity 36.1%; Pred. No. 3.7e-60;
 Matches 256; Conservative 122; Mismatches 251; Indels 81; Gaps 14;
 QY 49 SGEIVQELLSDPTDVELSGYSLTRDVGAARKGRVYQNDDEIRARYDKSVPLK 108
 DB 35 SHLDEVKRVMBEPRKTVKLGCELTITISQVAALIAARDNEVAVQLAESSRAGYKASDWM 94
 QY 109 AOLQNSV--YGVTTGSGSADTTEDAVSLQKALIEHOLCGVTPTSXSFVGRLENTL 166
 DB 95 ESMKKGTDSYGVTTGFGATSHRTKQKALQKELLRFNLNAGI-----FNGTGSTHTL 147
 QY 167 PLEVVGAMVIRVNSLTRGSAVRLVLEALNPLNHRITPIVPLRGSIASGDLSPASY 226
 DB 148 PHSATRAMLVRLINTLQSGYSGRFEILEITIKFLNHNITPCLPRGTTASGDLVPLSY 207
 QY 227 IAGATGHPVYKVAVLHEGTEKTFAREALISLFGLEAVV--LGRREGGLVNGTAVSAS 283
 DB 208 IAGLTGRPNKSA--VQPNQ--ESLNAEQAFKLAGNSGIFELQKKEGLAVNGTAVASG 264
 QY 284 MATSLSDHMLSLTSCALTALTVEMVQCGSPAFPHDVCRPHRGVEVARNRITLS 343
 DB 265 LASATLAEANITLLEWMSAVFAEVMNKP--EFDHDLTHKIKHPGIEAALVEHILD 323
 QY 344 GSSFAVEHEEVKVDDEGLRDYRYLRTSPQLPGLVEDMNAVSTLSLENTTTDNP 403
 DB 324 GSSYV-YAAQKMEHMDPLQPKKODRYALRTSPQMLQPGVEIRATKMIEREINSVNDP 382
 QY 404 LIDVENKQTAHGNRPQASAVSISMEKTRIALILKTNFTOTELNANAKRGIPSCLA 462
 DB 383 LIDVSRKALHGNFQGPPIGVSMDSRLAISIKLMPAQSELVNDFYNNGLPSNLISG 442
 QY 463 AEDPSLNYHGKGLDIIIAAYASELGHANPVTTFVQAPBMGNQAVNSIALISARTAEAN 522
 DB 443 GRNPSLDYGKGEIEMASVCELOFLANPVNHOQAEQHNQDNQNSGLISSKRTYBAL 502
 QY 523 DVLSLLASHLYTTLQAVDLRAMELDFKQFPDLLPTLLQOHLGTLGVNALAEVKKAL 582
 DB 503 DIHKMSSSTYVALCAQIDLRLHE-----ENIRLSVKNXI 537
 QY 583 NKRLQOTTYDLEPRHND-----FSY-----ATGTVA----- 611
 DB 538 SQVAKTILTTGVNGELHPSRCELDLIRVDBREYFAVYDPCSTTYLAKMLQVLEH 597
 QY 612 -LISSPSANVTTLAVANAKVASAKAISLIREVEN--RFQOTFSSQAPAAHAYLSPRTRVL 669
 DB 598 ALKNGENENKASTSISIQKIEAFEAELKAVLPKEVASAHVALEDDKPAIANRITCEGRSYPL 657
 QY 670 YSPVREELGVQARGDVPFYVQGEITIGSVNSIYALIDGRINIVLYML 719
 DB 658 YKIFREELGT-----NFLTEKVMSPGEBODVFAKSGLIVDPLKCL 702

DR 10-OCT-2003 (Rel. 42, last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Camellia sinensis (Tea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Ericales; Theaceae; Camellia.
 NCBI_TaxID=4442;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yabukita; TISSUE=Leaf;
 RA Matsumoto S.; Takeuchi A.; Hayatsu M.; Kondo S.;
 RT "Molecular cloning of phenylalanine ammonia-lyase cDNA and
 RT classification of varieties and cultivars of tea plants (Camellia
 RT sinensis) using the tea PAL cDNA probe."
 RL Theor. Appl. Genet. 89:671-675(1994).
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC - SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D26556; BAA05643.1; -.
 DR HSSP: P21310; 188F.
 DR InterPro: IPR008948; L-Asparagine-lyase.
 DR InterPro: IPR001106; Phe/His_NH3lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRPFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 200 202 5-imidazolinone (Ala-Gly)
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 714 AA; 77751 MW; 5F0E7EBD0A9071 CRC64;
 Query Match 28.0%; Score 1012; DB 1; Length 714;
 Best Local Similarity 36.2%; Pred. No. 4.7e-60;
 Matches 268; Conservative 130; Mismatches 280; Indels 62; Gaps 18;
 QY 5 LDSLATTANGFTNGSHAAPTSAAGPTSLARRPGLDGHAAGSQGEIVQELLSDPTD 64
 DB 1 MDS-TTIAINGV--GSGSGPEFCIKDPLNM-----GVAARMKGSILHEEVKGVVEERKP 52
 QY 65 VVELSGYSLTRDVGAARKGRVYQNDDEIRARYDKSVPLKAOQNSV--YGVTTG 122
 DB 53 VVRLQGEFTLITISQVAALIAVRGSEVAVELSSARBGVYASDWMESMNKGTDSYGVTTG 112
 QY 123 GGSADTTEDAVSLQKALIEHOLCGVTPTSXSFVGRGLE--NTPLEVVGAMVIRV 180
 DB 113 GATSHRTKQKALQKELLRFNLNAGI-----FNGTGSTHTPQSATRAMLVRLIN 163
 QY 181 SLTRGSAVRLVLEALTNPLNHRITPIVPLRGSIASGDLSPASYAGATGHPVYKV 240
 DB 164 TLLQSGYSGIRFEILEALISKPLNNITPCLPRGTTASGDLVPLSYIAGLTGRPNKSA-- 222
 QY 241 VLHEGTEKIMFARALISLFGLEA--VVLGRREGGLVNGTAVSASAVTTLSDHMLSL 298
 DB 223 --VQPTGELIHPKEAFRLAGVEGGEFELQKKEGLAVNGTAVASGLASVLEFANILAVL 280

```

QY 299 SQALTLALTEAVNGQGGSFAPPHIDVCPRPGQVAVANRITLLSSGFVNEHEEYKVK 358
DB 281 SEVLALIFAEVWQGP-BFTDHLTHKLHHPQGLPAALIMHILIDSSY-VKAAQKHEM 338
QY 359 DDEGILRODRVPLRTSPQFLGLVEDMMAVSTLLENNTTNDPLLDVENKQTAHGNF 418
DB 339 DLQKPKQORVLRTPKWLGLPIEIVISSSTKSIEREINSVNDPPLINVSNNKALHGNF 398
QY 419 QASAVSISNEKTRALLALIGKLNFTQCTELLMANNRGLPSCLA-AEDPSLNTHGKLDI 477
DB 399 QGPFGVSDMNRRLVAVSIGKLMFQPSBELVNDPFYNNGLPSNLGGRNPSLDYFKKAEI 458
QY 478 HIAAVASELGHANVTTEVQPAEMGNQAVNSIALISARPAEANDVLSLLASHLYCTL 537
DB 459 APMAYGSELQFLANVTNHVQSAEQHNDVNSLGILSSRKTAEDVILKIMSTIYVALC 518
QY 538 QAVDRLAMELDFKQFDPLLPILLQOHLGTGLDVNALALEYKALKNRLEQ----- 588
DB 519 QAVDLRHFENLRNTVSTVSQVAKRVLTMG--VNG-ELHPSRFCEKOLLRVDRERYFA 575
QY 589 -----TTTYDEPRKMDAFSAYTGVVELSSPSANVTLTAVNAMAFAEKAIS--L 640
DB 576 YIDDPGSAVTPMQRQ-----VIVEHALKNESEKKNLSTIFQKIRAFEEIKTLL 628
QY 641 TREVRN-RFWQTPSSQAPAHAYLSPTRTVLSFVFEELGVQARRGDYFVGVQOETIGSNV 699
DB 629 PREVESTRALENGNSAIPNRKKEGRSPLYKFFVREELGT-----ELLTGKVASPGEEF 683
QY 700 SRTYEAIKDGRIHVLYKML 719
DB 684 DKVFALCKGEMLDPLMCLT 703

RESULT 15
PALI_SOYBN STANDARD; PRT; 713 AA.
ID P27991.1
AC 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=cv. T225; TISSUE=leaf;
RC MEDLINE=92190550; PubMed=1799682;
RA Frank R.L., Vocklin L.O.;
RT "Sequence and structure of a phenylalanine ammonia-lyase gene from
RT Glycine max.";
RL DNA Seq. 1:335-346(1991).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52953; CAA37129.1; -.
DB FIR; S22991.; S22991.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Asparatase-like.
DR InterPro; IPR001106; phe/His_NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLNC 199 201
FT (By similarity).
FT DHA (2,3-DIDEHYDROALANINE)
FT MOD_RES 200 200
FT (BY SIMILARITY).
SQ SEQUENCE 713 AA; 77744 MW; 9D71EF1CC230216A CRC64;

Query Match 28.0%; Score 1011; DB 1; Length 713;
Best Local Similarity 36.9%; Pred. No. 5.5e-60;
Matches 270; Conservative 120; Mismatches 282; Indels 60; Gaps 19;

QY 14 NGFTGSHAAPTAKSAAGPTSLRRTPGLDGHAAHQSOLEIYQELLSPDTPDVELSEYSL 73
DB 5 NQHNGSFCGLST-AKGNNDPLNM--GAAAEAMGSHLDEYKRVAVAEKRVYALGGETL 60
QY 74 TVRDVGAARKGRVRYVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGSADTRTE 131
DB 61 TIAQYAAVAGHDGVAVALHESAREGVAKSSEMVMNSMNGTSDSYGVTGGCAISHRKX 120
QY 132 DAVSLQKALIEHQLCGYVPTKXSFSGVSGLENTLPLEVNGAVNIIVNSLTRGSAVRL 191
DB 121 QGGALQKELIRFLNAGI-----FGNGTSSHTLPHTATPAALVRLNTLLQGSIRF 173
QY 192 VYLEALTNPLNRIPPIVPLRSGISASGSLPSLYAGAITGHPDVKXVHLHEGTEKIMF 251
DB 174 ELLEBATKLNNNVPRCIDLKGTITASGDVPLSYIAGLIGRNSKA--VQPSG--EVLN 230
QY 252 ARBAISLFGEL--EAVVLGPKEGELGVNGTANVASAKNTLSLDSHMLSLISQALITALYEA 309
DB 231 AKAEAFELASINSEFELPKKEGLALVNGTAVGSGLASVWLEFANITLAVSLVAIFAEV 290
QY 310 MYGQGGSFAPPHIDVCPRPGQVEVARNIRITLLSGSPAVE---HEBEVAVKQDEGILR 365
DB 291 MQGKP-EFTDHLTHKLHHPQGLPAALIMHILIDSSY-SYMAKKLHELDPLQKP-----K 344
QY 366 QDRYPLRTSPQFLGLVEDMMAVSTLLENNTTNDPLLDVENKQTAHGNFQASAVSI 425
DB 345 QDRYALRTSPQWLGLPIEIVIRFSTKSIEREINSVNDPPLIDVSRNKALHGNFQGPPIGV 404
QY 426 SMEKTRLALALIGKLNFTQCTELLMANNRGLPSCLA-AEDPSLNTHGKGLDIHAAVAS 484
DB 405 SMDNTRILALASIGKLMFAQFSELVNDPFYNNGLPSNLTASRNPSLDYGFKGAEIYASICS 464
QY 485 ELGHLANPYTTFVQPAEMGNQAVNSIALISARPAEANDVLSLLASHLYCTLQAVDLRA 544
DB 465 ELQYLANPVTTHVQSAEQHNDVNSLGILSSRKTEAIEILKIMSTIYVALCQIDLRH 524
QY 545 MEIDFKQGFDPPLLPILLQOHLGTGLDVNALALEYKALKNLKLEQTTTDLBRPHMDASY 604
DB 525 LSENLKNSVNTVVSQVSKRILTTG--VNG-ELHPSRFCEKOLLRVDRERY-----ISY 575
QY 605 ---ATGVVELLSPPSANVTLTAVNA-----MKVASAEKAIS--LTREVRNRF 648
DB 576 IDPPGSAVTPMQRQKQVLDVLAALVAECKQVNSIIFQKLAIFEEELKNLLPKVEGAR 635
QY 649 WQTPSSQAPAHAYLSR-RTRVLYSFVFEELGVQARRGDYFVGVQOETIGSNVRIYEAIK 707
DB 636 AAVESGKAALPKKIQCRSVPYKFFVREELGT-----GLLTGKVASPGEEFKLFTAMC 690
QY 708 DGRINHLYKML 719
DB 691 QGRKIIDPLMECL 702

```

Mon Sep 13 10:31:12 2004

us-09-939-408a-13.rsp

Page 14

Search completed: September 9, 2004, 10:04:39
Job time : 19.4315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:57:28 / Search time 87.6349 Seconds
(without alignments)
2592.269 Million cell updates/sec

Title: US-09-939-408a-13
Perfect score: 3610
Sequence: 1 MAPSIDSIATLTIANGFTNGS.....RIYEAKIDGRINHVLVLMGLA 720

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1041.5	28.9	725	10 Q94ANI	Q94ANI arabidopsis
2	1031.5	28.6	730	10 Q9W567	Q9W567 rubus idaeu
3	1025	28.4	711	10 Q8W2E4	Q8W2E4 lactuca sat
4	1018	28.2	711	10 Q9AXI5	Q9AXI5 pharbitilis n
5	1016.5	28.2	707	10 Q7X7G3	Q7X7G3 oryza sativ
6	1016.5	28.2	717	10 Q8RWP4	Q8RWP4 arabidopsis
7	1014.5	28.1	708	10 Q8RWP4	Q8RWP4 rehmmania g
8	1013	28.1	715	10 Q852S1	Q852S1 daucus caro
9	1010.5	28.0	714	10 Q8H6V6	Q8H6V6 populus tre
10	1007.5	27.9	714	10 Q7X6M8	Q7X6M8 oryza sativ
11	1003.5	27.8	671	10 Q84VE0	Q84VE0 oryza sativ
12	1002.5	27.8	721	10 Q9XFX5	Q9XFX5 citrus clem
13	998	27.6	713	10 Q8GZR8	Q8GZR8 lactuca sat
14	997.5	27.6	703	10 Q8YXG7	Q8YXG7 zea mays (m
15	997	27.6	716	10 Q9ATN7	Q9ATN7 asaracae r
16	994	27.5	717	10 Q8H6W0	Q8H6W0 coffea cane

17	991.5	27.5	716	10 Q9MAX1	Q9MAX1 catharanthu
18	990.5	27.4	710	10 Q94C45	Q94C45 manihot esc
19	990	27.4	715	10 Q24266	Q24266 populus kit
20	984.5	27.3	687	10 Q94EP0	Q94EP0 manihot esc
21	982	27.2	711	10 Q8H6V5	Q8H6V5 populus tre
22	980.5	27.2	712	10 Q94F83	Q94F83 manihot esc
23	971.5	26.9	655	10 Q9XGR3	Q9XGR3 vigna ungui
24	967.5	26.8	718	10 Q9XFX6	Q9XFX6 citrus clem
25	957.5	26.5	703	10 Q84LI2	Q84LI2 phalaenopsi
26	957	26.5	681	10 Q8RUZ3	Q8RUZ3 pinus sylve
27	956	26.5	681	10 Q8SAS6	Q8SAS6 pinus sylve
28	956	26.5	681	10 Q8RV49	Q8RV49 pinus sylve
29	955	26.5	549	10 Q04876	Q04876 hordeum vul
30	950	26.3	619	10 Q8SAS7	Q8SAS7 pinus sylve
31	946	26.2	681	10 Q39545	Q39545 cucumis mel
32	936	25.9	619	10 Q8H6V9	Q8H6V9 coffea cane
33	934	25.9	618	10 Q9FS00	Q9FS00 dianthus ca
34	930.5	25.8	727	10 Q7XJC3	Q7XJC3 pinus pinas
35	929.5	25.7	727	10 Q7XJC4	Q7XJC4 pinus pinas
36	786	21.8	396	10 Q8VWP6	Q8VWP6 beta vulgar
37	780.5	21.6	532	10 Q81E47	Q81E47 capsicum ch
38	769.5	21.3	544	10 Q9FT38	Q9FT38 rubus idaeu
39	741.5	20.5	362	10 Q9FUI7	Q9FUI7 ginkgo bilo
40	741.5	20.5	380	10 Q84U27	Q84U27 rubus idaeu
41	714.5	19.8	363	10 Q9FUI8	Q9FUI8 hordeum vul
42	659.5	18.3	497	10 Q04869	Q04869 hordeum vul
43	653.5	18.2	475	10 Q04875	Q04875 hordeum vul
44	620	17.2	435	10 Q9MSU1	Q9MSU1 betula verr
45	619	17.1	436	10 Q9LEP1	Q9LEP1 betula verr

ALIGNMENTS

RESULT 1
ID Q94ANI PRELIMINARY; PRT; 725 AA.
AC Q94ANI;
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia lyase (Hypothetical protein) (BC 4.3.1.5).
GN A12G37040 OR PAL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koeseima E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinozaki K., Shinozaki K., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.,
RT "Full length cDNA of gene T1J8-22/At2G37040 (GI:6598547).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamita A.,
RA Kawai J., Kim C., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Shinozaki K., Shinozaki K., Tang C.C., Toriumi M.,
RA Davis R.W.,
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Ergu P., Lee J.M.,

RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RY
 RP [4]
 RP SEQUENCE FROM N.A.
 RA Cochran F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.,
 RT "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
 RT Characterization of Recombinant Proteins and Knock-out Mutant
 RT Characterization";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY045919; AK76593.1; -;
 DR EMBL; BT003330; AA029949.1; -;
 DR EMBL; AY079363; AAL85094.1; -;
 DR EMBL; AY03128; AAP59438.1; -;
 DR PIR; G84787; G84787.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia ligase activity; IEA.
 DR GO; GO:0016841; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Hypothetical protein; Lyase.
 SQ SEQUENCE 725 AA; 78725 MW; 026263B2DDE9CE CRC64;
 Query Match 28.9%; Score 1041.5; DB 10; Length 725;
 Best Local Similarity 36.5%; Pred. No. 3.2e-61;
 Matches 272; Conservative 138; Mismatches 258; Indels 77; Gaps 21;
 QY 18 NGSHAAPTSAGPTSL-----RTPGIDGHAHQ---SOLEIYQELLSD 60
 DB 4 NGAH---KSNGGVAMLCGDIDIKKNVINAEDPLMGAALAEQMSKSHLDEKRWAE 59
 QY 61 PTDDVVELSGYSLTRVDVVGARKGRVRVQNDDEIRARVDXVDLFAQLQNSV--YGV 118
 DB 60 FRKPVNLTGETLTICQVAISTIGNSVYELSETARAGVNASDWMVSNKKTDSYGV 119
 QY 119 TTGFGSADRTEDAVSLQALIEHQLCVPTKSFSVGRGLENLPLEVVGAMVIR 178
 DB 120 TTGFGATSHRTKNGVALQKELIRPLNGI-----RSTKETSHLPHSATPAALVR 172
 QY 179 VNSLTGHSANRVLVLEALTNFLNRHRTPIVRLGSGISASGDLPSYIAGATGHPDVX 238
 DB 173 INTLGGFSGIRFEIEATISFLNNITPSPLELRTITASGDVLPLSTIAGLLGRPNRK 232
 QY 239 VHTLHGTE-KIMFAEALSLFGLA--VLGPKKGLGVNCTRAVSAMATLSHDSHML 295
 DB 233 A---TGNGEALTAEAKLGISSGFEDLPQKGLLVNGTAGSGMAVMVLFETNVL 288
 QY 296 SLISQALTLTVEAMVQCGQSFAPFIHVCRRHPQOVAVANIRITLSSGSAFVHEEEV 355
 DB 289 SYLAETLSAVFAEVNSGKP-EFTDHLTRLKHHPQIEAALIMEHIDGSSY-WKLAQKL 346
 QY 356 KVVQDEGILRODRYPIRTSPQLGPLVEDMMAHASTLSLENNTTTNDPLDVENKQYAG 415
 DB 347 HEMDPLQKPKQRYALRTSPQWLGGQIEVIRATKSIEREINSVNDPLDIVSRKAHIG 406
 QY 416 GNFQASAVSISMEKTELATLIGKLNFTQCTELNNAANRGJPSCL-AAEDPSLNYGKG 474
 DB 407 GNFQGFPIGVSMNDTRLALIGKLMFQFSELVNDVFNNGLPSNLNRSRPSLDYFGKG 466
 QY 475 LDIHIAVASELGHLANPYTTPVQPAENGQAVNSLALISARTPAANDVLSLLASHLY 534
 DB 467 AEAAMASYSELCQYLANPVTSHVQSAEONQDVSLSISRKTSSEAVDILKIMSTFLV 526
 QY 535 CTLQAVDLRAMELDFKQDPLPLTLQCHLGTLGVNALLAELVKKALNKL-----EQ 588

DB 527 AICQAVDLHLHLENIRQIVKNTVSQVAKKVLITG--VNG-ELHSPRFCEKOLLKYVDREQ 583
 QY 589 TTTYDEPRMHAFASTACTIVE-----LSSSPSANVLTIANMAKVASAKAI 638
 DB 584 VYTYADP-----CSATYPLIQKLVQIVDHALNGSEKNAVTSIFH--KIGAFEEEL 635
 QY 639 S--LTREVA--NRFWQTSQAPAHAYLSPRTVLYSFVEELGVQARQDPFVGOQET 694
 DB 636 KAVLPREVAAQAADNGNSAIP-NRIECSYPLRYRREELGT-----ELLTGEKYS 689
 QY 695 IGSNVSRIYEAIKQGRINHLVYKML 719
 DB 690 PGEFEDKVTATICEGKIIDPMWECI 714
 RESULT 2
 ID Q9MS67 PRELIMINARY; PRT; 730 AA.
 AC Q9MS67;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
 GN PAL2.
 OS Rubus idaeus (Raspberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Rosaceae; Rosaceae; Rosoideae; Rubus.
 OX NCBI_TaxID=32247;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21437948; PubMed=11553751;
 RA Kumar A., Ellis B.E.;
 RT "The phenylalanine ammonia-lyase gene family in raspberry, structure,
 RT expression, and evolution";
 RL Plant Physiol. 127:230-239(2001).
 DR EMBL; AP237955; AAP40224.1; -;
 DR HSBP; P21310; 1B8F.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia ligase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 730 AA; 79422 MW; 0E942417B50CE113 CRC64;
 Query Match 28.6%; Score 1031.5; DB 10; Length 730;
 Best Local Similarity 37.1%; Pred. No. 1.5e-60;
 Matches 268; Conservative 130; Mismatches 276; Indels 49; Gaps 17;
 QY 14 NGFTNGS--HAAPKSAAGPTSLARTRPGIDGHAHQ---SOLEIYQELLSDPTDVVEL 68
 DB 13 NGIQGSLDDGLCIKTSIKTGYSDPLMVGAAABSGMGSILDEVRAVAAEYRKPVNL 72
 QY 69 SGYSLTVDVVGAAKGRVRVQNDDEIRARVDXVDLFAQLQNSV--YGVTTGFGGSA 126
 DB 73 GGETITISQVAALINHGSGVAVELAEARAGVASSDWMVSNKKTDSYGVYTTGFGATS 132
 QY 127 DTRTEDAVSLQALIEHQLCVPTKSFSVGRGLENLPLEVVGAMVIRVNSLTRGH 186
 DB 133 HRTKQGAALQKELIRPLNAGVLANGTES-----AHTLPHSATPAALVINTLLQGY 185
 QY 187 SAVRLVLEALTNFLNRHRTPIVRLGSGISASGDLPSYIAGATGHPDVGVHVEGT 246
 DB 186 SGIRFEIEALSKLNINITPCLPLRGTITASGDVLPLSTIAGLLTGRPNRKA-VQPKG- 243

QY 247 EKIMAREARISLFGLEA--VVLGPKREGILVNGTAVASAMATSLHDSHMLSLLSQALTA 304
 Db 244 -ETLNAAEFAVAGISSGPFELQPREGALANGVAGSLASTYLFEFTNLALSELISA 302
 QY 305 LITVAVVGGGSGFAFPIHDVCRPHGQVEVAVARNTRLTSSGSPFAVEHEEVKXVDDEGL 364
 Db 303 IFAEVMGQKP-EFTDHLTHKLKHHPGQLEAALMEHILDGSSY-VKAAEKLEQDPLQKP 360
 QY 365 RODRPLRTSPPLCPPLVEDMMKHAIVTSLNNTTTNPLLDVENKQTAGNFGQASAVS 424
 Db 361 KODRYALRTSPWLPQIEVIRFSTKSTIERINSVNDPLIDVRNRLGQGNFGQTPIG 420
 QY 425 ISMEKTRIALALIGKLNFTQCTELNANMNGLPSCLA-AEDPSLNYHGKGLDHIHAAVA 483
 Db 421 VSMQDTRIALAISIGLMAFQPSBELVNDVYNGGLSPNSLSCGRDPLDGFPGAEIAMSASYC 480
 QY 484 SELGHLANPTTTFVQPAENGQAVNSLALISARRTAENDVLSLLASHLYCTLOAVDLR 543
 Db 481 SELGLANPTTNTVHVSQAEGHNDVNSLGLISSRKTAEAVIDILKLMSSTFVALCOAIDLR 540
 QY 544 ANELDPKQGFDDLLFTLLQCHIGTGL-----DVNALALEVKALNKRLSEQ--TTT 591
 Db 541 HLEENLKSTVKNVTQOLARVLTTVNGELHPSRCEKDLNVVERELPRAYIDPESAT 600
 QY 592 YDLERPMHDAFSYATGTVVELLSSSPSANVTLLTAVNAWKVASAKAIS--LTREVRN-RF 648
 Db 601 YLMQRLRGVL-----VEHALTNGENENKMASTSI-FQXITAFEEBELKTLIPKEVESARA 653
 QY 649 WOTPSSQAPAHAYLSPTRTVLVYSPFREELGVQARGDVFGVQOETIGSNVRIYEAIKD 708
 Db 654 AYESGAALPNRIIVECRSYPLKPFVEBELG-----GEPLTEKVRSPREECQKFTAMCQ 708
 QY 709 GRI 711
 Db 709 GRI 711
 RESULT 3
 Q8W2E4
 ID 08W2E4 PRELIMINARY; PRT; 711 AA.
 AC 08W2E4;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase.
 OS Lactuca sativa (Garden lettuce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
 OC Lactuca.
 OC NCBI_TaxID=4236;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RA Campos-Vargas R., Nonogaki H., Suslow T., Saltveit M.;
 RT "Characterization of phenylalanine ammonia-lyase (PAL) gene in wounded
 lettuce leaf tissue".
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DB EMBL; AF299330; AAL5542.1; -.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; P:ammonia-lyase activity; IEA.
 DR GO; GO:0016841; P:ammonia-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro: IPR008948; L-Asparatase-like.
 DR InterPro: IPR001106; Phe/His NHlyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 711 AA; 77363 MW; 3120AB2877C29513 CRC64;

Query Match 28.4%; Score 1025; DB 10; Length 711;
 Best Local Similarity 37.9%; Pred. No. 4e-60;
 Matches 267; Conservative 127; Mismatches 259; Indels 52; Gaps 18;
 QY 40 GUDGAHAGQLEIVQELSDPDVVELSGSYLTVRDVVG--AARKGRVYRQNDDEIR 97
 Db 23 GVAAEALGSHLDEKXKVAEFRKPVVKLGSELTIVSQVAGIAANDSDTVKYLESEAR 82
 QY 98 ARVKSUVFLAQQLQNSV--YGVTFGFGSADPTREDAVSLQCALIHQCGVPTSSXS 155
 Db 83 AGVAASSVWVESNKKGTDSYGVTTGFGATSHRTKGGALQKELLIFLNAI----- 135
 QY 156 FSVGRLGENTLPLEVGAIVRVNSLTRGSAVRLVLEALTNFNLHRTTPIVPLRGSI 215
 Db 136 FONGTSEHTLPHSATRAAMIVRINTLLQYSGIRFEILAIETFLNNNTTTPCLPIRGTI 195
 QY 216 SASGLSLSTIAGALNCPDVKXVHLHEGTEKIMFEREALISLFGLEA--VVLGPKREG 273
 Db 136 TASGLVLSYIAGLIGTRPNSKA--VGPTGVLANEKAPAAAGVGGFFELQPREGLA 252
 QY 274 LVNGAVASAMATSLHDSHMLSLLSQALTTALITVAVVGGGSAFPIHDVCRPHGQVE 333
 Db 253 LVNGTAVSGMAAMVLPFANLALLSEVLSAIFAEVMGQKP-EFTDHLTHKLKHHPQIE 311
 QY 334 VARNITRLSSGSPFAVEHEEVKXVDDEGLIRQDRYPLRTSPPLGSLVEDMMKHAIVTSL 393
 Db 312 AAALMEYILDGSDY-VKAAQRYHEMDPLQKPKQDRYALRTSPWLPQIEVISTYMI 370
 QY 394 IENNTTNPPLDVENKQTAGNFGQASAVISMEKTRIALALIGKLNFTQCTELNANM 453
 Db 371 REINSVNDPLIDVSRKALHGSNFGCTPIGVSDNTRRLAALIGKLMFQFSLVVDVY 430
 QY 454 NRGDPSCLA-AEDPSLNYHGKGLDHIHAAVSELGHLANPTTTFVQPAENGQAVNSLAL 512
 Db 431 NNGLPNSLGGNRPFLDYGFKGSIAMASYCELOFLANPTTNTVHVSQAEGHNDVNSLGL 490
 QY 513 ISARRTAENDVLSLLASHLYCTLOAVDLRAMELDPKQGFDDLLFTLLQCHIGTGLDVN 572
 Db 491 ISARRTAENDVLSLLASHLYCTLOAVDLRAMELDPKQGFDDLLFTLLQCHIGTGLDVN 548
 QY 573 ALALEVKALNKRLSEQTTTVDLERPMHDAFSY--ATGT-----VVELLSSSP 617
 Db 549 G-ELIPSRFCEKDLRVVDREY-----VFAYIDVCGSTYPLMQKRVLDHALNNGE 601
 QY 618 SANVTLLTAVNAWKVASAKAIS--LTREVRN-RFWQTPSSQAPAHAYLSPTRTVLVYEFVR 674
 Db 602 TEKNTNTSI-FQXIATFEEBELKVLIPKEVEGRVAYENDTLISNRIKACRSYPLRYEVR 660
 QY 675 EELGVQARGDVFGVQOETIGSNVRIYEAIKGRINHVLVKKL 719
 Db 661 BELG-----RG-FLTGEKVTSPGEHEDRVFTAMCKQIIDLPLEGL 700
 RESULT 4
 Q9AXI5
 ID 09AXI5 PRELIMINARY; PRT; 711 AA.
 AC 09AXI5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 OS Phacelis nil (Violet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamids; Solanales; Corvulaceae; Ipomoea.
 OC NCBI_TaxID=35883;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RA STRAIN=cv. Violet;
 RA Nakazawa A., Nozue K., Yasuda H., Takeba G., Kubo H.;
 RT "Expression of phenylalanine ammonia-lyase in Phacelis nil";
 Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.


```

Db      576 ANPLMTKIRAVLVEHALANGPAEKDSSVHS-KITAFEEELREALPREMEARVAFT 634
Qy      654 SQAP-AHAYLSPTRYLVSFVEELGVQARRGDVFGVQOETIGSNVRIEAIKDRIN 712
Db      635 GNAPINRKRKERSFPLVYFVAREELGV-----YLTGKLSPGEECKXVLAISEKRLI 689
Qy      713 HVLVKKL 719
Db      690 DPMLECL 696

RESULT 6
O8RMP4 PRELIMINARY; PRT; 717 AA.
ID O8RMP4 AC
O8RMP4 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase.
AT3053260.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxId=3702;
OX
RN
[1]
SEQUENCE FROM N.A.
RP Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RP Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RP Palm C.J., Bowser L., Jones T., Banh U., Gerninci P., Chen H.,
RP Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RP Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RP Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shimozaki K.,
RP Ecker J., Theologis A., Davis R.W.,
RP Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN
[2]
SEQUENCE FROM N.A.
RP Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RP Palm C.J., Bowser L., Jones T., Banh U., Gerninci P., Chen H.,
RP Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RP Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RP Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shimozaki K.,
RP Ecker J., Theologis A., Davis R.W.,
RP Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN
EMBL: BT000035; AAM12956.1; -
DR GO: GO:0005737; C-cytoplasm; IEA.
DR GO: GO:0016211; F-ammonia lyase activity; IEA.
DR GO: GO:0016841; F-ammonia-lyase activity; IEA.
DR GO: GO:0016829; F-lyase activity; IEA.
DR GO: GO:0009058; P-biosynthesis; IEA.
DR GO: GO:0006559; P-phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Aspartase-like.
DR InterPro: IPR001106; Phe/His NH3-lyase.
DR InterPro: IPR005922; Phe-am-lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRPFAMs: TIGR01226; phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KM
LYase.
SQ SEQUENCE 717 AA; 77801 MW; C4C9CBDCBDB91A838 CRC64;

Query Match 28.2%; Score 1016.5; DB 10; Length 717;
Best Local Similarity 36.0%; Pred. No. 1.5e-59;
Matches 264; Conservative 141; Mismatches 283; Indels 45; Gaps 17;
Db      1 MDQIEMLCGGSGKTVAVTTKLADPLNW-----GLADQKSHDEVKMAAEIRRP 55
Qy      65 VVELSGVSLVRDVGARKGRVRVONDEIRARVDSVDFLKAQLONSV--YGVTTGF 122
Db      56 VVNLGSELTILIGVAIAISTVGGVKVELAETSRAGVYKASDVMVMSMKKGTDSYVTTGF 115
Qy      123 GGSADRTEDAVSLQALLIEHQLCGVTPISXSSFSVSGLENTLPLEVVRGAMVIRNVL 182

```

```

Db      116 GATSHRRRTKNGTALOTELIRPLNAGI-----FGNTKETCHTLPQATRAMLVAVNTL 168
Qy      183 TGHSAVRLVLEALTNPLNRITPIVPLRGSISASGDLSPISYAGATGHPVXVHL 242
Db      169 LOGYGINFEILEATISLNNNISPSPLRGTITASGDLVPLSLIAGLIRPSKA--- 225
Qy      243 HEGTE-KIMFARFALSFLGLA--VLGPKGGLGVNGTAVASAMATLSLSDHMSLLS 299
Db      226 -TGPDESLTAKAEAEKAGISTGFEDLPQKEGLLVNGTAVAGSNMAVLFENAVQAVLA 284
Qy      300 QALTALTYEAMVQGGSPAPFPIHDVCRPHPGQVEVARNIRITLSSGSPAHEEVEVKVD 359
Db      285 EYLSAIFAEVNSGR-EFTDHITRLKHPQCIKAAIMEILLGSSY-MKLAQKHEMD 342
Qy      360 DEGLIKDRPYLRTSPQELGPEVEDMMAVSTSLNNTTNDPLDVENKQTAHGNFQ 419
Db      343 PLOKROFVALRTSPQELGPEVEDMMAVSTSLNNTTNDPLDVENKQTAHGNFQ 402
Qy      420 ASAVSISMEKTRIALALGKLNFTQCTELLNAAANRGLPGL-AAEPPSLNHHKGLDIH 478
Db      403 GPPIGSMNDRLALIAAGKMFQAFSSLVNDFYNNGLPSVLTASNPISLDYGRKGAELA 462
Qy      479 IAAVASELGHLANPYTTTQVPAEMGNQAVNSLALISARTAEANDVLSLASHLYCTLQ 538
Db      463 MASYSSELQYLANPVTSHVQAGAECHNDVNSLGLISSKISEAVDILKMTSTFLVGICQ 522
Qy      539 AYDLRAMELDFKKQDPLPTLQOHLGTGLDVNALAEVKALANKL-----EQTTY 592
Db      523 AYDLRAMELDFKKQDPLPTLQOHLGTGLDVNALAEVKALANKL-----EQTTY 579
Qy      593 DLEP--RWMDASVATGTVVE-LSSPSANVTLLAVANMVASAEKAS-LTREVNR 647
Db      580 VDDPSAITYPLMQRROYIVDHALSNGETEKNAVTSI-FQKIGAFEEIKKAVLKVEYEA 638
Qy      648 FQWTSQAPAHAYLSP-RIRVLYSFVEELGVQARRGDVFGVQOETIGSNVRIEAI 706
Db      639 RAAYNGTAPIPNRKEXRSPVLYFVAREELGKLTIGEXV-----SPGEEFDFVTAM 693
Qy      707 KQGRINHLVKKL 719
Db      694 CEKGLDPLMDCL 706

RESULT 7
O94ENO PRELIMINARY; PRT; 708 AA.
ID O94ENO AC
O94ENO 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL1
OS Rehmannia glutinosa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Lamiales incertae sedis; Rehmannia.
OX NCBI_TaxId=99300;
RN
[1]
SEQUENCE FROM N.A.
RP STRAIN=CV. Keumtan;
RA Yun S.U., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RT "Molecular characterization of phenylalanine ammonia-lyase gene in
RT Rehmannia glutinosa.";
RN
EMBL: AF401636; AAK84225.1; -
DR GO: GO:0005737; C-cytoplasm; IEA.
DR GO: GO:0016211; F-ammonia lyase activity; IEA.
DR GO: GO:0016829; F-lyase activity; IEA.
DR GO: GO:004548; F-phenylalanine ammonia-lyase activity; IEA.
DR GO: GO:0009058; P-biosynthesis; IEA.
DR GO: GO:0006559; P-phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Aspartase-like.

```

DR InterPro: IPR001106; Phe/His NHase.
DR InterPro: IPR005922; Phe am lyase.
DR InterPro: IPR006162; Prantase_S.
DR Pfam: PF00221; PAL_1.
DR TIGRfams: TIGR01226; phe.am.lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDINASE; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
KW Lyase.
SEQUENCE 708 AA; 76835 MW; 1B6AF08816F1B8FE CRC64;

Query Match 28.1%; Score 1014.5; DB 10; Length 708;
Best Local Similarity 35.6%; Pred. No. 2e-59;
Matches 264; Conservative 122; Mismatches 269; Indels 87; Gaps 17;

DR 18 NGSHAAP---TKSAGPTSLRRTPGLDGHAAHQSLEIVQELSDPTDDVVELSGYSLT 74
DB 3 NHHHNSGLCVELTRPPLMNVAAESLKG-----SHLDVKKMVEEFKRPVKLGESLT 57
QY 75 VRDYVGAKRGRRVRYQNDDEIRARYDKSVDFLKAQLONSV--YGVTTGGGSGADRTED 132
DB 58 IAQVAALAAADNAVAVELAETAPAGVKASSDWMESEMKGTDSYGVTTGGFATSHRTKQ 117
QY 133 AVSLQKALIEHQCGVTPTSXSFVSGRGLENTLPLEVVRGAMVIRVNSLGRGSAVLY 192
DB 118 GGLQKELIFPLNAGT-----FGNGTESNFALEHSAATRAAMLVRIITLLQGSIGIFFE 170
QY 193 VLEALTNFNRITPPIVPLRGISASAGDLSPLSYIAGALTGHDPVKVHLHGTEKIMFA 252
DB 171 ILBALTFELNHNITPCLPRGTTASGDVLPYSIAGLLTGRPSNKA-VGPNQ--EALNA 227
QY 253 REAISI.FGLEAVY-LGPKGGLGVNCTASASAMATLSLHDSNLSLSQALTALTVEMV 311
DB 228 GEAFSLAGVSGFELPKKGLALVNGTAVSGSLASIALYDANILVALISEVTSVFAEVMN 287
QY 312 GQGSFAPFIHDVCRPHQGVAVARNIRITLSSGFAYHEEVEVYKDDGILRQDRYPL 371
DB 288 GKR-EFTDHLTKLKHHPQIDEMALMEHILDGSAY-VKAAQKHEPTPLPKQPDORAL 345
QY 372 RTRPQFLGLVDMMAAYSTLSLENTTTDNLPLEVENKQTAHGNFQASAVSISMEXTK 431
DB 346 RTRPQWLGQIEVIRATKIEREINSVNDPLDIVSRNKAHGNFQGTPIGVSMNTR 405
QY 432 LAALIGKINFTQCTELNLAAMNRLPSCIA-AEDPSLNYHGKGLDIIHAYASELGHIA 490
DB 406 LAIAAIGKLMFAQFSLVNDFYNNGLPSNLSGRRPSLDYGRKGEIEMASCYSELQFLA 465
QY 491 NPVTTVQPAEMGNQAVNSLALISARTRAEANDVLSLLASHLYCTLOAVDLAMELDFK 550
DB 466 NPVTNHVQSAECHNQVNSLIGLISRKVEALDIKLMSSTYLALCQAVDLRHLDE---- 521
QY 551 KQDPLPLPTLQGHHTGLDVNAALAEVKKALKRLEQTTTYDDEPRHDA----- 601
DB 522 -----ENLRISVKNIVSQVAKKILTNGLNGELHPSRFEKDLR 560
QY 602 ---FSY---ATGT-----VVELLSSPS-ANVTLTAVANAKVASAKAI 638
DB 561 VVDREYVFAYIDDPGCTYPLMGKLRQVLVDHALNNGSEKKNVTSIQKLEAREVELKA 620
QY 639 SLTRERAN-RFWGTBSSQAPALAYLSPRTRVLYSVREBELGVQARGGVFVQOQETIGS 697
DB 621 ILRKEVASARIALAESGNPAIGNRTECRSYPLKFRBELGTNYLTGKVV-----SPGE 675
QY 698 NVSRIVEAIKGRINHLVKKML 719
DB 676 ECDKVFALSKGLIVDPLKCL 697

RESULT 8
Q85251
ID Q85251 PRELIMINARY; PRT; 715 AA.
AC Q85251;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN GPCPAL3.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulid; Apiales; Apiaceae; Apioidae; Scandiceae; Daucinae;
OC Daucus.
OC NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Kurodagosun;
RA Ozeki Y., Chikagawa Y., Kimura S., Soh H., Maeda K., Pornsiriwong W.,
RA Kato M., Akimoto H., Oyanagi M., Fukuda T., Koda T., Itoh Y.,
RA Yamada A., Ueno H., Takeda J.,
RT "Putative cis-elements in the promoter region of phenylalanine
RT ammonia-lyase gene of carrot involved in anthocyanin biosynthetic
RT pathway";
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB069813; BAC56977.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0005058; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0005559; P:biosynthesis; IEA.
DR InterPro: IPR0018948; L-Asparagine-lyase.
DR InterPro: IPR001106; Phe/His NHase.
DR InterPro: IPR005922; Phe am lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRfams: TIGR01226; phe.am.lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDINASE; 1.
KW Lyase.
SEQUENCE 715 AA; 77802 MW; 8D06779D5D7565B4 CRC64;

Query Match 28.1%; Score 1013; DB 10; Length 715;
Best Local Similarity 36.1%; Pred. No. 2.6e-59;
Matches 265; Conservative 131; Mismatches 275; Indels 64; Gaps 18;

QY 16 FTNGSHAAPTSAAGPTSLRRTP-----GLDGHAAHQSLEIVQELSDPTDDVVELSGY 71
DB 3 YTNHGH-----ENGNGVLDLCKMKEDPLSMGVAAELKSHHEEVRVAAEVRKRVVYIKGG 58
QY 72 SLTVRDVVG-AARKGRVRVQNDDEIRARYDKSVDFLKAQLONSV--YGVTTGGGSGADT 128
DB 59 TLISQVAAISARDDSGVKVELSEAAAGVKASSDWMESEMKGTDSYGVTTGGFATSHR 118
QY 129 RTEDAVSLQKALIEHQCGVTPTSXSFVSGRGLE--NTLPLEVVRGAMVIRVNSLITRG 185
DB 119 RTKQGALQKELIFPLNAGI-----FGSGAEAGNNTLPHSATRAALVAINTLLQ 169
QY 166 HSAVALVLEALTNFNRITPPIVPLRGISASAGDLSPLSYIAGALTGHDPVKVHLHGTE 245
DB 170 YSGIRFELIEATFELNHNITPCLPRGTTASGDVLPYSIAGLLTGRPSNKA--VGP 226
QY 246 TEKIMFAEASLSFLEA--VVLGPKGGLGVNCTASASAMATLSLHDSNLSLSQALT 303
DB 227 TGVTLSPPEARFKLAGVBSGFELPKKGLALVNGTAVSGSAMVLEFANILVALEVMNS 286
QY 304 ALTVFAMTGGQGSFAPFIHDVCRPHQGVAVARNIRITLSSGFAYHEEVEVYKDDGIL 363
DB 287 ALFAEVMQGR-EFTDHLTKLKHHPQIDEMALMEHILDGSAY-VKAAQKHEKDPLOK 344
QY 364 LEODRYPTSPQFLGLVDMMAAYSTLSLENTTTDNLPLEVENKQTAHGNFQASAV 423
DB 345 PKQDRYARTSPQWLGQIEVIRSTYKIEREINSVNDPLDIVSRNKAHGNFQGTPI 404
QY 424 SISKEKTLALALIGKINFTQCTELNLAAMNRLPSCIA-AEDPSLNYHGKGLDIIHAY 482
DB 405 GVSXDNTRLALAAIGKLMFAQFSLVNDFYNNGLPSNLSGRRPSLDYGRKGEIEMAS 464
QY 483 ASEIGHLANPVTTVQPAEMGNQAVNSLALISARTRAEANDVLSLLASHLYCTLOAVDL 542

```

Db      465 CSELOFLANPVNHNQSAEQHNQVNSGLISRSKTSRAVEILKMTSTFLVGLCOAIDL 524
QY      543 RAMEIDFKKQEDPELLPTLLQOHLGTGLDVNALALEVKKALKRLEQ-----568
Db      525 RHLEENLKSTVANTVSQVAKRYLTWG--VNG-ELHPSRFCEKDLRAVDREYIFAYIDDP 581
QY      589 -TTTTDLERHMDASVYATGTVVELLSSPSANVLTAVNWKVASAKAIS--LTREVR 645
Db      582 GSATTPLOKRE-----TLVEHALNNGDKERNLSTISFOKIAFEDELKALLPREVE 634
QY      646 N-RFWQTSQAPAHAYLSPRTVLVSFVREELGVQARRGDVFVQOETIGSNVSIYE 704
Db      635 SARAAVESGNPAIPRIKRCRSYPLKVFREELGT-----EYLGEKVTSPGEEDKVF 689
QY      705 AIKDGRIHVLYKML 719
Db      690 AMTKGELIDPLECL 704

RESULT 9
Q8H6V6 PRELIMINARY; PRT; 714 AA.
AC      Q8H6V6;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Phenylalanine ammonia-lyase.
GN      PAL1.
OS      Populus tremuloides (Quaking aspen).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosid1; Malpighiales; Salicaceae; Salicaceae; Populus.
OX      NCBI_TaxID=3693;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22264006; PubMed=12376645;
RA      Kao Y.Y., Harding S.A., Tsai C.J.;
RT      "Differential expression of two distinct phenylalanine ammonia-lyase
RT      genes in condensed tannin-accumulating and lignifying cells of quaking
RT      aspen."
RL      Plant Physiol. 130:796-807(2002).
DR      EMBL; AF480619; FAN52279.1; -.
DR      GO; GO:0005737; Cytoplasm; IEA.
DR      GO; GO:0016211; Gamma-ammonia ligase activity; IEA.
DR      GO; GO:0016841; Gamma-ammonia-lyase activity; IEA.
DR      GO; GO:0016829; Gamma-ammonia-lyase activity; IEA.
DR      GO; GO:0009058; P-biosynthesis; IEA.
DR      GO; GO:0006559; P-phenylalanine catabolism; IEA.
DR      InterPro; IPR008948; L-asparatase-like.
DR      InterPro; IPR001106; Phe_His_NH3lyase.
DR      InterPro; IPR005922; Phe_am_lyase.
DR      Pfam; PF00221; PAL; 1.
DR      TIGRFAM6; TIGR01226; Phe_am_lyase; 1.
DR      PROSITE; PS00488; PAL_HISTIDASE; 1.
KM      lyase.
SQ      SEQUENCE 714 AA; 77556 MW; 9BDADA21CECC35D CRC64;

Query Match      28.0%; Score 1010.5; DB 10; Length 714;
Best Local Similarity 35.5%; Pred. No. 3.8e-59;
Matches 265; Conservative 135; Mismatches 269; Indels 77; Gaps 21;

QY      8 LATTLANGTNGSHAFTSAAGPTSAARRTP---GLDGAHSGOLEIYVELLSDPTDD 64
Db      1 METITNGVQNGS-----SESLCTGRDPLSGVAAEAMKSHLDEVKRMVAERKP 51
QY      65 VVELSGYSITLVDPVVG-AARKGRRVVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTG 121
Db      52 VVWLAQGNLTIAQVASIAGHDASNVVELSEBAPRVKASSPWVMSDKDTSYGVTGG 111
QY      122 FGGSATRTREDASVLOKALIEHQLCGVTPTSKSSSVGRGLE--NTLPEVVRGAMVIRV 179
Db      112 FGATSHRRTKQGALQKELIRFLNAGI-----FNGTGETCHTLPHSATRAAMLVRI 162

```

```

QY      180 NSLRGSAVRVLVYLEALTNFNHRITPIVPLRGSIASGDLSPISYIAGATGHPVKV 239
Db      163 NTLQGISGIRFELIEALITKLNANNITPCLPKGTIASGDVPLSTIAGLINSRPSKA 222
QY      240 HVLHGTE-KIMFARERAIISLGEA--VILGPKGLGVNGTAVSASMATLSLHDSHLS 296
Db      223 ---TGPGVEVLDAVEAFKAGIDSGFEELQPKGLNLVNGTAVGSLAMVLFETVLA 278
QY      297 LLSQALTLTLEAVNGOOGSFAPIHDVCRPHQGVAVANNITLLSGSPAYE---HE 352
Db      279 VLSLISALFPAEVNNGR-EFTDHLTKLKHPEQILPAALIMEHITLDGSAYMRAAKLH 337
QY      353 EEVKYQDEGLRDRYPLRTSPQFLGPIVEDMMAHAYSTLSLENNITTDNPLDVENKOT 412
Db      338 MDPLQKP-----KQDRVALRTSPQWLGPGQIEVIRFSKTSEREINSVNDPRLDVSRNKA 392
QY      413 AHGNGFQASVANSISNEKRLALIGKLNFTQCELLNAAMNGRLPSCL-AAADPSLNTH 471
Db      393 LHGNFQGTPIGVSNDVYRLALIASIGKLLFAQFELNDYFNNGLPSNLPAASNPISLDYG 452
QY      472 GKGLDHTAAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAANDVLSLLAS 531
Db      453 FKGALIMASYSCELOFLANPVNHNQSAEQHNQVNSGLISRSKTAFAVDILKLMSTT 512
QY      532 HLVTQLQAVDIRAMEIDFKKQEDPELLPTLLQOHLGTGLDVNALALEVKKALKRLEQTTT 591
Db      513 FLVALCOAIDIRHLEENLRSAVKNVSHVSRVLTGANB--LHPSRCEKELLKVV- 568
QY      592 YDLERHMDASVYATG-----TYVELSSPSANVLTAVNWKVASAK 636
Db      569 -DRE---DVRAVADDPESATYPLMQKRLQVLYHALANGENKASATSV-FQKIAFED 622
QY      637 AIS--LTREVRN-RFWQTSQAPAHAYLSPRTVLVSFVREELGVQARRGDVFVQOQ 693
Db      623 ELKALLPREVESARAAYDSGNSALENKIKECRSYPLKVFREELGT-----GLITGENVR 677
QY      694 TIGSNVSRITYEALNDGRINHVLYKML 719
Db      678 SPGEEDFVFTAMCGKXIDPLECL 703

RESULT 10
Q7X6M8 PRELIMINARY; PRT; 714 AA.
ID      Q7X6M8;
AC      Q7X6M8;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      OSUNBA0073R02.16 protein (OSUNB0061C13.5 protein).
GN      OSUNBA0073R02.16 OR OSUNB0061C13.5.
OS      Oryza sativa (Rice).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCBI_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA      Liu Y.L., Wu J., Yu Z., Chen L., Fan D.L., Meng Q.J., Zhang L.,
RA      Lu Y.Q., Yu S.L., Liu S.H., Lu T.T., Zhang Y.J., Li Y., Li C., Li T.,
RA      Zhang Y., Hu H., Jia P.X., Qian Y.M., Yang K., Zhou B., Chen Z.H.,
RA      Hao P., Zhang L., Mu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA      Ren S.X., Lv G., Lin W., Gu W.Q., Zhang J.F., Tu Y.F., Jia J., Yin H.F.,
RA      Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA      Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA      Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL      Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AL731616; CAE05456.1; -.
DR      EMBL; AL731629; CAE05623.1; -.
SQ      SEQUENCE 714 AA; 76937 MW; C4C21FF6E0D588FA CRC64;

Query Match      27.9%; Score 1007.5; DB 10; Length 714;
Best Local Similarity 36.4%; Pred. No. 6e-59;

```



```
Q9XFX5
ID Q9XFX5 PRELIMINARY; PRT; 721 AA.
OC Q9XFX5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phenylalanine-ammonia lyase (EC 4.3.1.5).
GN PAL1.
OS Citrus clementina x Citrus reticulata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=93374;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Fortune; TISSUE=Flavedo;
RC Sanchez-Ballester M.T., Lafuente M.T., Zacarias L., Granell A.;
RT "Phenylalanine ammonia-lyase gene expression in response to low
RT temperature in mandarin fruit";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AA238753; CABA42793.1; -.
DR HSSP; P21310; 1B8F.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfams; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW lyase.
SQ
SEQUENCE 721 AA; 78429 MW; 970BD6351ASFDA5 CRC64;
Query Match 27.8%; Score 1002.5; DB 10; Length 721;
Best Local Similarity 37.4%; Pred. No. 1.3e-58;
Matches 262; Conservative 124; Mismatches 253; Indels 61; Gaps 18;
QY 49 SQLEIVQELSDPTDVEVLSGLTVRDVVGAAKKG--RRVYQNDDEIRAVYKXVD 105
DB SHLEEKVAAVYRRPVVNLGGETLTVAAVAIAGVNVAAQVVELSESAKEGVKXSSD 102
QY 106 FLKAQLQNSV--YGVTTGFGSADTRTEDAVSLQKALIEHQLCGVTFSXSFVGKGLE 163
DB 103 WWDSDNKKITDSYGVTTGFGATSHRTQNGALQKELIKFANAGI-----FANGTKSS 155
QY 164 NTLPLEVVGAMVIRVNSLTRGSAVRLVLEALTNFLNRITPIVPLRGSISASGDLSP 223
DB 156 HTLPHSATRAAMLVHVNITLLQYSGIRFEIILDAITKLINHSITPCLPLRGITTSAGDLVP 215
QY 224 LVIYAGATIGHDVAKVHVEHETE-KIMFARPAISLPLEAVNLGPKREGGLVNVGTVA 282
DB 216 LSTIAQLTLGRPNRKA---TSPNGEIIIDAGASQAQFGFPELOPEGLAVLVGTAVGS 271
QY 283 SVATSLHDSHMLSLISQALITALTEAWVGGQGSFAPFIHDCVPHPGQVAVANIRITLL 342
DB 272 GLASVWLPANNLALISELISAIPEAVQKRP-ETDHLTLKHKHPQIQIAAAMHITL 330
QY 343 SGGSFVVEHEEVKVKDDEGILIRQRIPLRTSPQPLGLPLVDVMAHASTLSLENNITTDN 402
DB 331 DGSSY-VKAAKLTHEIDPLQKPKQDRYALRTSPQPLQIIVIRFATKSIEREINSVND 389
QY 403 PLIDVENKOTAFAGNFOASAVSISMEKRETLALATGKLNFGCELTLPAAANRGLPSCLA 462
DB 390 PLIDVENKALHAGNFQGTPIGVSMNDHRLAALGKLMFAQFSLVNDFFNNLPSNLS 449
QY 463 -AEDPSLNTHGGLDTHIAVYSEIHLANPVTTVQPAEWGQAVNSLALISARPAEA 521
DB 450 GGRNPSLIDYGFAGAEIAYASCSLEQLFANPVTNVSQAEQHNODVNSIGLISRKTAEA 509
```

```
QY 522 NDVLSLLASHLYCTLQAVDIRAMELDFKQFDPPLPTLLQOHLGTLDVNALAEVKA 581
DB 510 VDLIKLMSTFLVNLQCAIDRLHLEENLKHIVKDVQVQARXKVLTVANGB---LHPSRF 566
QY 582 LNKRLQCTTVDLEPRWDASY-----ATGTVELL-----SSSPSAVTL 623
DB 567 CKKDLKXA-----DREA-VFAYIDDPCSATYPLMQRLQVYVHALNENENGNASSI 620
QY 624 TAVNAKVASAEKALIS--LTREVRNRFWQTPSSQAPA--HAYISPRTRVLYSFVEELGV 679
DB 621 FQ-----KIAAEHEELKAVLPKEVENA-RQYVENGNPTIPRIKECRSGYPLRYLVEELGT 675
QY 680 QARQGVFVQVQOENIGSNVSRIVEATIKDGRINNVLYKML 719
DB 676 ----NFLTGEKVTSPGEKFDKFTAMQCKITIDPMTECL 710
RESULT 13
Q8GZR8 PRELIMINARY; PRT; 713 AA.
AC Q8GZR8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase2.
OS Lactuca sativa (garden lettuce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Asterales; Asterales; Cichoriaceae; Cichoriaceae;
OC Lactuca.
OX NCBI_TaxID=4236;
RN [1]
RP SEQUENCE FROM N.A.
RA Campos-Vargas R., Nonogaki H., Suslow T., Salveit M.E.;
RT "Characterization of a second phenylalanine ammonia-lyase (PAL) gene
RT in wounded lettuce leaf tissue";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411134; AA013347.1; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfams; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW lyase.
SQ
SEQUENCE 713 AA; 77657 MW; 87751A2ED7349E61 CRC64;
Query Match 27.6%; Score 998; DB 10; Length 713;
Best Local Similarity 36.8%; Pred. No. 2.6e-58;
Matches 256; Conservative 122; Mismatches 264; Indels 54; Gaps 16;
QY 40 GUDGHAHQSQLEIVQELSDPTDVELSGSLTVRDVVG-AAKGRRRVQNDDEIRA 98
DB 29 GMAASLNGSHLDEVKRVVAFRRKPVVRLGGETLTVGVAIAASDNAGYVELSEFARA 88
QY 99 RVDKXVDPLKAQLQNSV--YGVTTGFGSADTRTEDAVSLQKALIEHQLCGVTFSXSF 156
DB 89 GVKASSDWMSNMKGTDSYGVTTGFGATSHRTKEGALQKELIRLNGI-----F 141
QY 157 SVGRGIENTLPLEVVGAMVIRVNSLTRGSAVRLVLEALTNFLNRITPIVPLRGSIS 216
DB 142 NGGTSTHTLPHSARAAMLVRINTLLQYSGIRFEIILATKLNINNVDPPLRGITIT 201
QY 217 ASGDISPLSYIAGATGDPDVYHVLHEGTEKIMFARPAISLPLEA--VYLQKXELGL 274
DB 202 ASGDIVPLSYIAGLTGRANRKA---VQPTGEVNAKRAFAEAGVBOGFPELOKEXIAL 258
```



```

RT   rugosa";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AR326116; AKK15640.1; -.
DR   HSSP; F21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0005559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L:aspartase-like.
DR   InterPro; IPR01106; Phe/Hs_MH3lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRPFAM; TIGR01226; phe am lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KM   Lyase.
SQ   SEQUENCE   716 AA; 77487 MW; 9903F33B8CFD2958 CRC64;

Query Match      27.6%; Score 997; DB 10; Length 716;
Best Local Similarity 35.2%; Pred. No. 3,1e-58;
Matches 262; Conservative 126; Mismatches 271; Indels 86; Gaps 16;

QY   13  ANGFTNGSHAAATKGAAGTSAIRTPGLDGHAAHQSLIVQELSDPTDDVIELSGYS 72
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    9  SNGHNNGANGFCVKQ-NDPLWMAAAESLKG----SHLEEVKRMVEEFRKPVVKLGGET 62

QY   73  LTVRDVVGARKGRVRYONDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRT 130
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    63  LTISQVAAIAAKDNVAVELAESARAAGYKASSDWMDSKGTDSYGTTGFGATSHRT 122

QY   131 EDVLSIQALLIEHQICGVTPTSXSSFSVGRGLENTPLEVVGAMVIRVNSLTRGSAVR 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    123 KGGGALQKELLRFNLNAGIFGNATES-----NHTLPHTATRAMLVRINTLLQYSGIR 175

QY   191 LVVLALTNPLNHRITPTVPLRGSSASGDSPLSYTAGITGHPDYKHYVHGEOTEKIM 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    176 FEILALITFELNQNTPCPLRGTITASGDLVPLSYTAGLITGRPNKA-VGRAG--EPL 232

QY   251 FAREALISLGLER--VLGPEKGLVNGTAVSASMATLSLHDSHMLSLSQALTALTVE 308
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    233 TAEQAFKLAGVTCGFEDLPKEGLALVNGTAVSGSLASIALFDANVLAVLSVMSPVFAE 292

QY   309 AMVGGQGSAPFIHVCRPHGQVAVANIRTLISGSSFAVHEEEYKVKDDEGILRQDR 368
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    293 VNNGRP-EFTDHLTKLKHHPQGLEAAIMEHILDGSGY-VKAAQKLHEIDPLQPKQDR 350

QY   369 YPLRTSPQFLGLVEDMMHAYSTLSLENNTTDNPLLDVENKQTAHGNFQASAVSISYE 428
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    351 VALRTSPQWLGQIENVITATKCIEREINSVNDPLDIVSRNKAIHGNGFGTPIGYSMD 410

QY   429 KTRLLALIGKLNFTQCTELLNAANNRGLPSCLA-AEDPSLNYHGKGLDHIHAAVASELG 487
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    411 NTRLLIASIGKLLFQAQFSELVNDPFNNGLPSNLGGRNPSLDYGFKGSRIAMASYSCELSQ 470

QY   488 HLANPYTTFVOPAEKNGQAVNSLALISARTAFANDVLSLLASHLYCTLOAVDLRAMEL 547
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    471 FLVNVPTNVHQAEOHQDNQVNSLGLISRKTYEALDIDIKLMSSTYVALCOAVDLRVEE 530

QY   548 DEKQDFDPLPTLLIQOHGTGLDVNALALEVKKALNKRLEQTITTYDLEPRWDA----- 601
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    531 NKK-----LAVKNTVSQVAKRTITMGVNGELHPSRFCEKE 565

QY   602 -----FSY-----ATGTVE-----LSSSPSANVTLTAVNAKVASAE 635
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    566 LIRVVDREYVFYIIDPCLATPYLMQKLRQVLVDHALKNGSEKENVSTSIPIKTEAPEEE 625

QY   636 KAISLTREVRN-RFWQTPSSCAPAHAYLSPTRVLYSFVREELGVQARRGDVFWGQOET 694
    ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
    626 LKALLPKVESARIALBSGSPAVANRIEGRSPLYKFIREEELGTGLTGEXAV-----S 680

QY   695 IGSNVSRIEAIKGRINNVLYKYL 719
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

DB 681 PGECEKVFPAALNSGLIIDPLELCTL 705

Search completed: September 9, 2004, 10:07:42
Job time : 91.6349 secs

(without alignments)
7040.653 Million cell updates/sec

Sequence: 1 atggccccctcctcgactc.....nnnnnnnnnnnnnnnnnnn 2475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

Database :

1:	gb	ba:	*
2:	gb	hg:	*
3:	gb	in:	*
4:	gd	om:	*
5:	gd	ov:	*
6:	gb	pac:	*
7:	gb	ph:	*
8:	gb	pl:	*
9:	gb	pr:	*
10:	gd	ro:	*
11:	gb	secs:	*
12:	gb	sy:	*
13:	gb	un:	*
14:	gb	vi:	*
15:	em	ba:	*
16:	em	fun:	*
17:	em	hum:	*
18:	em	in:	*
19:	em	ma:	*
20:	em	om:	*
21:	em	or:	*
22:	em	ov:	*
23:	em	pat:	*
24:	em	ph:	*
25:	em	pl:	*
26:	em	ro:	*
27:	em	secs:	*
28:	em	un:	*
29:	em	vi:	*
30:	em	hg	hum:
31:	em	hg	inv:
32:	em	hg	other:
33:	em	hg	mus:
34:	em	hg	pin:
35:	em	hg	rod:
36:	em	hg	mam:
37:	em	hg	vrt:
38:	em	sv:	*
39:	em	hgo	hum:
40:	em	hgo	mus:
41:	em	hgo	other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

[illegible]

ALIGNMENTS

RESULT 1			
AR199606			
LOCUS	AR199606	2475 bp	DNA
DEFINITION	Sequence 20 from patent US 6355468.		linear
ACCESSION	AR199606		
VERSION	AR199606.1	GI:20249680	
KEYWORDS	.		
SOURCE	unknown.		
ORGANISM	unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2475)		
TITLE	Yoshida,R.K. and Koetsura,A.B. Phenylalanine ammonia lyase polypeptide and polynucleotide sequences and methods of obtaining and using same		
JOURNAL	Patent: US 6355468-A 20 12-MAR-2002;		

QY 2044 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCGGCGAGCTCTTCTCGGCAAGCAGAGG 2100
|
|
|
Db 2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCGGCGAGCTCTTCTCGGCAAGCAGAGG 2100
|
|
|
QY 2101 TGAAGTGGGCAACCAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 2160
|
|
|
Db 2101 TGAAGTGGGCAACCAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 2160
|
|
|
QY 2161 AGCTCTCTGCAAGATGCTCGCTAGNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 2220
|
|
|
Db 2161 AGCTCTCTGCAAGATGCTCGCTAGNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 2220
|
|
|
QY 2221 NNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNN 2280
|
|
|
Db 2221 NNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNN 2280
|
|
|
QY 2281 TNNNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 2340
|
|
|
Db 2281 TNNNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 2340
|
|
|
QY 2341 NGTNNCCANNNAACNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNN 2400
|
|
|
Db 2341 NGTNNCCANNNAACNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNN 2400
|
|
|
QY 2401 NNNANAAA 2408
|
|
|
Db 2401 NNNANAAA 2408
|
|
|
RESULT 2
AX366868 2475 bp DNA linear PAT 15-FEB-2002
LOCUS
DEFINITION Sequence 20 from Patent W00208402.
ACCESSION AX366868
VERSION AX366868.1 GI:18698157
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.
REFERENCE
1 Yoshida,R.K. and Koopstra,A.B.
Phenylalanine ammonia lyase polypeptide and polynucleotide
sequences and methods of obtaining and using same
Patent: WO 0208402-A 20 31-JAN-2002;
JOURNAL
PCBU Services, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2475
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Consensus Sequence of SEQ ID Nos: 12, 15, and 18"
ORIGIN
Query Match 86.0%; Score 2128; DB 6; Length 2475;
Best Local Similarity 100.0%; Pred.No. 7,4e-224;
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTCGTACGCTGCGAGCTCGAGCTCGGCAACGCGTCCGCAAGGTTNG 60
|
|
|
Db 1 ATGGCCCTCCCTCGTACGCTGCGAGCTCGAGCTCGGCAACGCGTCCGCAAGGTTNG 60
|
|
|
QY 61 CAGCGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 120
|
|
|
Db 61 CAGCGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 120
|
|
|
QY 121 GCTGCTCTCTCCGCAACCAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCA 180
|
|
|
Db 121 GCTGCTCTCTCCGCAACCAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCA 180
|
|
|
QY 181 ACCCGACCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 240
|
|
|
Db 181 ACCCGACCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 240
|
|
|

QY 241 TCGGCGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 300
|
|
|
Db 241 TCGGCGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 300
|
|
|
QY 301 AAGATCGAACAAGAGTGAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 360
|
|
|
Db 301 AAGATCGAACAAGAGTGAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 360
|
|
|
QY 361 AAGATCGAACAAGAGTGAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 420
|
|
|
Db 361 AAGATCGAACAAGAGTGAAGTCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 420
|
|
|
QY 421 GNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 480
|
|
|
Db 421 GNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 480
|
|
|
QY 481 GNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 540
|
|
|
Db 481 GNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 540
|
|
|
QY 541 GTCAACTGCTCAACGAGGATCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGG 600
|
|
|
Db 541 GTCAACTGCTCAACGAGGATCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGG 600
|
|
|
QY 601 AACTTCTCAACGAGGATCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGG 660
|
|
|
Db 601 AACTTCTCAACGAGGATCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGG 660
|
|
|
QY 661 GCGGACCTCTCCGCGGCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAAC 720
|
|
|
Db 661 GCGGACCTCTCCGCGGCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAAC 720
|
|
|
QY 721 GTNCAAGTNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 780
|
|
|
Db 721 GTNCAAGTNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 780
|
|
|
QY 781 TNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 840
|
|
|
Db 781 TNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 840
|
|
|
QY 841 GCGGACCTCTCCGCGGCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAAC 900
|
|
|
Db 841 GCGGACCTCTCCGCGGCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAAC 900
|
|
|
QY 901 TCGGAGGCGGCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAG 960
|
|
|
Db 901 TCGGAGGCGGCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAG 960
|
|
|
QY 961 CCGTCTCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGTCAACG 1020
|
|
|
Db 961 CCGTCTCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGTCAACG 1020
|
|
|
QY 1021 CCGGAGGCTCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGT 1080
|
|
|
Db 1021 CCGGAGGCTCTCAACGAGTCAACCCCATGCTCCGCGCGGCAACGAGTCAACGAGT 1080
|
|
|
QY 1081 GAGGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1140
|
|
|
Db 1081 GAGGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1140
|
|
|
QY 1141 GAGGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1200
|
|
|
Db 1141 GAGGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1200
|
|
|
QY 1201 TGAAGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1260
|
|
|
Db 1201 TGAAGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1260
|
|
|
QY 1261 TGAAGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1320
|
|
|
Db 1261 TGAAGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1320
|
|
|
QY 1321 GCGGAGGAGGATCTCCGCACTCAAGAGCCATCAAGNCGGCGCATCAACC 1380
|
|
|

Db 1321 GGGAGGCTCACTTCAAGAGCTCAGAGATGCTCAAGCGCGGATGATACCGCGGCTT 1380
1381 CCGTCCGCTGCTGCTGCGAGGAGCCGCTGCTCTCTATCACTGCAAGGCTCGACATT 1440
1381 CCGTCCGCTGCTGCTGCGAGGAGCCGCTGCTCTCTATCACTGCAAGGCTCGACATT 1440
1441 GCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1441 GCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1501 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1501 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1561 ACNGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1561 ACNGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1621 CAGCGCGTGCATCTCCGCGGATGAGTTCAGATTCAAGAGCAGTTCGACCCGATNTTC 1680
1621 CAGCGCGTGCATCTCCGCGGATGAGTTCAGATTCAAGAGCAGTTCGACCCGATNTTC 1680
1681 NCNCGCTGCTCAAGAGGAGCTTGGGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGG 1740
1681 NCNCGCTGCTCAAGAGGAGCTTGGGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGG 1740
1741 CTGCGGAGCAAGGTCACCAAGCGCTCNAAGCGCTGAGGAGCACTGCTGAGC 1800
1741 CTGCGGAGCAAGGTCACCAAGCGCTCNAAGCGCTGAGGAGCACTGCTGAGC 1800
1801 CTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1801 CTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1861 NNGTCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1861 NNGTCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1921 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1921 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1981 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
1981 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
2041 TCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
2041 TCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
2101 TCGCGATCGGCAAGAGCTCTCCGCGATCAAGAGGCACTCAAGGCGGCGATCAACC 2160
2101 TCGCGATCGGCAAGAGCTCTCCGCGATCAAGAGGCACTCAAGGCGGCGATCAACC 2160
2161 ACCTCTCTGCTCAAGATGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
2161 ACCTCTCTGCTCAAGATGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
2221 NNNCCNNNNNNNNCTTTTNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
2221 NNNCCNNNNNNNNCTTTTNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
2281 TNNNNCTNNCTNNCTCNCNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
2281 TNNNNCTNNCTNNCTCNCNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
2341 NCTNNNNNNNNNNCTNNNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
2341 NCTNNNNNNNNNNCTNNNNNNNNNNCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
2401 NNNNNAAA 2408
2401 NNNNNAAA 2408

Db 2401 NNNNNAAA 2408

RESULT 3
AR199605
LOCUS AR199605 18 from patent US 6355468. 2439 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence
ACCESSION AR199605
VERSION AR199605.1 GI:20249679
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2439)
Yoshida, R.K. and Koobstra, A.B.
Phenylalanine ammonia lyase polypeptide and polynucleotide
sequences and methods of obtaining and using same
Patent: US 6355468-A 18-12-MAR-2002;
JOURNAL
FEATURES
source location/Qualifiers
1..2439
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 68.7%; Score 1699.4; DB 6; Length 2439;
Best Local Similarity 79.4%; Pred. No. 5.1e-177;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

1 ATGGCCCCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
1 ATGGCCCCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
61 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
61 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
61 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
61 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
121 GCT 180
121 GCT 180
101 GCT 160
101 GCT 160
181 ACCCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
181 ACCCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
161 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 217
161 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 217
241 TCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
241 TCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
218 TCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 276
218 TCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 276
301 AAGATCGACCAAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
301 AAGATCGACCAAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
277 AAGATCGACCAAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 336
277 AAGATCGACCAAGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 336
361 ACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
361 ACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
337 ACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 386
337 ACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 386
421 GCGCT 480
421 GCGCT 480
397 GCT 456
397 GCT 456
481 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
481 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
457 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
457 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
541 GTCAGCT 600
541 GTCAGCT 600
517 GTCAGCT 576
517 GTCAGCT 576
601 AACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
601 AACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
577 AACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 636
577 AACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 636
661 GCGAGCT 720
661 GCGAGCT 720

[illegible]

```
/db_xref="GI:18698156"
/db_xref="REMBRMBL:CAD23831"
translation="MAPSLDLSHSPANVAKAQAQVNGASTNLAVGSHLPTQVTO
VDIVERMLAAPTOSTLELDGYSLNTGDVVSARKR PVRVKDSDEIRKIDKVEFLR
SOLMSVYGVTTGFGSADRTEDIASLOKALLEHQLGVLPSFSFRLGRJENSL
PLEVRGAMTIRVNSLTRGHSARVLVLEALTNPLNHTPIVPLRGTISAGDLSPL
SYIAAISGHPDSKVHVHEGKILYAEEMALFNLEPVYLPKEGIGLVANGTAVSA
SMATLADHMHMILSLOSITAMTVYAWGAGHFRPHLHYTRHPQIIVANITIK
LLEGRFPAVHHEEYKVDDEGILRODRYPLRTSPQWJGPLVSDILHNAVLTEAQ
STTDNPLIDVENKTSHHGNFQAAAVANTMEXTRILGLOIGKLNFTQLEMLNGMR
GLPSCIAEDPISLHYKQJIDIAAAYISELGHANPLTTHVOPAEMLQAVNSLAI
SARRTESNDVLSILLATHLVCLQAIIDIRALEFPKQFGPAIVSLIDQHGSAVNA
SNRDELVEVKNITLAKRLQONSVDLVPRMHDASPAAGTVVEVLSSTSLAANA
KRYAAEASISLTRYRETFMFASTSSPALSYLSPROIILYAFRELGVKARGBV
FLQKQEVITIGSNVSKYEAIKSGRINNVILKRLA"
```

Query Match 68.7%; Score 1699.4; DB 6; Length 2439;

Best Local Similarity 79.4%; Pred. No. 5.1e-177;

Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

```
QY 1 ATGGCCCCCTCCCTCCAGCTGATCGGCACTCGNTGCGCAAGCGTCTCNMAAGGATNG 60
DB 1 ATGGACCCCTCGCTCGACTGATCTGCACTGGTGGCAAGGGGCTGCACTCCGCAAG 60
QY 61 CAGCGCCGNNCCGNNNGNCGNACGGGGCCAGCTCCACNCTCGNCGCCGCGNG 120
DB 61 CAGGCTGT-----CAATGGCGCTCGACCAACTC-----GCAATCGGAG 100
QY 121 GCTGCTCTCTCCGACCAACCAGNNGACCAAGTGAATCGTNGANAGATCCTCGCG 180
DB 101 GCTGCGACCTGCCCAACCAAGTGAATCGTNGANAGATCCTCGCGCG 160
QY 181 ACCCCACCGNACGNAAGNNTGAACTCGAGGGGTAACCTTCACTCCGAGGAGCGTG 240
DB 161 CGCGGACCGACTGCA---CGCTGAACTCGAGGCTACTCGTTCACCTCGGAGACGTG 217
QY 241 TCGGCGCGCGNCGCAAGGGCGCGNCGTCCGCTCNAGACGAGAGATCCGCGCA 300
DB 218 TCTCGCGCGGAGAGAGGCGAGCGCTGTCGCGTCAAGAGACAG--GAGAGATCCGCTCA 276
QY 301 AAGATGACAAANAGNCTGAGTTCTCCGNNCGNACGCTCNMAAGAGTCTAGAGNCTC 360
DB 277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTGCAACTCTTCATGAGGCTACGCGCTC 336
QY 361 ACGACTGATTTCCGCGGCTCGGCGACACCCGAGTGAAGATGCAATCTCGCTCAGAG 420
DB 337 ACGACTGATTTGGCGGATCCGGAGACCCGCAACGAGACGGCATCTGCTCCAGAG 396
QY 421 GGNCTCTCGAGACCAAGCTTGGCGGTGNTCTCCNAGTGAATGANTCTTCCGCTC 480
DB 397 GCTCTCTCGAGACCAAGCTTGGCGGTGNTCTCCGCTGCACTCTTCTCCGCTC 456
QY 481 GAGCGGAGCTCGAGAACTCGCTTCGAGTGAAGTCTCGGCGGCGCATACCATCGG 540
DB 457 GGGCGGGGCTCGAGAACTCGCTTCGCTCGAGGTTGTTCCGGGCGCATACCAATCGG 516
QY 541 GTCACTCGCTCAACGCGGCACTTCGCGNCTCGCTGCTGCTTCGAGGCGCTAC 600
DB 517 GTCAACAGCTTGACCGCGGCACTTCGCGTTCGCTGCTGCTGAGAGCGCTAC 576
QY 601 AACTTCTCAACAGCGGCACTACCCCATCGTCCGCTCCGCGGAGCATCTCGGCGTG 660
DB 577 AACTTCTCAACAGCGGCACTACCCCATCGTCCGCTCCGCGGAGCATCTCGGCGTG 636
QY 661 GGGAGCTTCTCCGCTCTCTACATCGCGCGCGCATCAACGCTCAACCGGACNNAG 720
DB 637 GGGAGCTTCTCTCTCTCTCTACATCGAGCGCGCATCAAGGGTCAACCGGACAG 696
QY 721 GTNCAAGTNNTCAGAGGCGANNAGAAATCATGTCGCGCGGAGGAGATGCGCTC 780
DB 697 GTGCAAGTCTGTCAGAGGCAAGAGAAATCTGTACGCGCGGAGGAGATGCGCTC 756
```

```
QY 781 TTNGGCTCGAGCCCGCTGCTCTCGGCGCGAGAGGCTCTGGTCTGTCACAGGACG 840
DB 757 TTCACCTCGAGCCCGCTGCTCTCGGCGCGAGAGGCTCTGGTCTGTCACAGGACG 816
QY 841 GCGCTCTCGGCTCGATGAGCGACCTCGCTCTGACGACGACCAAGTCTCTGCTCTC 900
DB 817 GCGCTCTGAGCATTCATGCGACCTCGCTCTGACGACGACCAAGTCTCTGCTCTC 876
QY 901 TCGCAGCGGCTCAACGCTCTTACGCTGAGGCGCATGCTCGGCGCGCTGCTTCC 960
DB 877 TCGCAGTCTGCTCAACGCTCTTACGCTGAGGCGCATGCTCGGCGCGCTGCTTCC 936
QY 961 CCNTTCTCTCAAGAGTCAAGCGCGCTGACCGGCGCGGAGATCGAGGCTCGGCGCAATC 1020
DB 937 CCTTCTCTTCAAGAGTCAAGCGCGCTGACCGGCGCGGAGATCGAGGCTCGGCGCAATC 996
QY 1021 CGCAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
DB 997 CGCAAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
QY 1081 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
DB 1057 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
QY 1141 GGGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB 1117 GGGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
DB 1177 TCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
QY 1261 TTCAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1237 TTCAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
QY 1321 GGCAGGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
DB 1297 GGCAGGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
QY 1381 CCNTTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
DB 1357 CCTTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 1441 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
DB 1417 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
QY 1501 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
DB 1477 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
QY 1561 ACGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
DB 1537 ACGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
QY 1621 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
DB 1597 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
QY 1681 NCGNCGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
DB 1657 GCTTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
QY 1741 CTGNGAGCAAGGTCACCAAGNCTGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
DB 1717 CTGNGAGCAAGGTCACCAAGNCTGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
QY 1801 CTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
DB 1777 CTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
QY 1861 NNGTCTCTGCGNNNGCANNAGGTCGCTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 1920
```

Db 1837 TCGTGCAGCTC-----GCTCTGCTCGCGCGCTCAACGCTTGAAGGTGCGCG 1885
 QY 1921 CCGCCGAGAGAGCCATCTTCGCTACACGCCANAGTCCGGAACCTTCTGCGCGCCGT 1980
 Db 1886 CCGCCGAGTCCGCTCTCTGCTACACGCCCAAGTCCGGAACCTTCTGCTGCGCGCGT 1945
 QY 1981 CCGTCCGCGCGCGCGCTCAGTCTCTGCTCTGCGCGCGCGCGCTCTGCTGCTGCTG 2040
 Db 1946 CCACTCTGCTGCGCGCGCTCTGCTCTGCTCTGCGCGCGCGCGCGCTCTGCTGCTG 2005
 QY 2041 TCCGCGAGAGAGCTCGCGCTCAAGGCGCGCGCGCGCGCGCTCTGCTGCGCGAGAG 2100
 Db 2006 TCCGCGAGAGAGCTCGCGCTCAAGGCGCGCGCGCGCGCGCTCTGCTGCGCGAGAG 2065
 QY 2101 TACAGATCGCGCGCGCGCGCTCTGCTCTGCTCTGCGCGCGCGCGCGCTCTGCTGCTG 2160
 Db 2066 TACAGATCGCGCGCGCGCGCTCTGCTCTGCTCTGCGCGCGCGCGCGCTCTGCTGCTG 2125
 QY 2161 ACCTCTCTGCTCAAGATGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2220
 Db 2126 AGTCTCTCTCAAGATGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2185
 QY 2221 NNNCCNN 2280
 Db 2186 TATCCCGCTGCACTCTTGAAGACTGCTCTTCTGCTGCACTGCTGCACTGCTCTT 2245
 QY 2281 TNNNNCTNNCTNNCTNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
 Db 2246 TCGTCTTGGCTGCT 2304
 QY 2341 NCTNN 2400
 Db 2305 AGTACCATTCAGAGCTCGATTAATCAAGGAGAGATCTAGCTGCGGAGCTCTCTG 2364
 QY 2401 NNNNNAAA 2408
 Db 2365 CGCATAAA 2372

RESULT 5
 E01783 2331 bp DNA linear PAT 29-SEP-1997
 Locus
 DEFINITION DNA sequence coding for L-phenylalanine ammonialyase.
 ACCESSION E01783
 VERSION E01783.1 GI:2170036
 KEYWORDS JP 1988317086-A/1.
 SOURCE Rhodospiridium toruloides
 ORGANISM Rhodospiridium toruloides
 Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
 Microbotrymycetidae; Sporidiobolales; Rhodospiridium.
 1 (bases 1 to 2331)
 Fukuhara,N., Yoshino,S., Matarabe,M., Nakajima,Y. and Makiguchi,N.
 RECOMBINANT PLASMID FOR MANIFESTING L-PHENYLALANINE-AMMONIA-LYASE
 AND TRANSFORMANT CONTAINING SAID PLASMID
 Patent: JP 1988317086-A 1 26-DEC-1988;
 MITSUBI TOAITSU CHEM INC
 JOURNAL
 COMMENT
 OS Rhodospiridium toruloides
 PN JP 1988317086-A/1
 PD 26-DEC-1988
 PF 18-JUN-1987 JP 1987152357
 PI FUKUHARA NOBUHIRO, YOSHINO SADAOKI, WATANABE MIDORI, PI
 NAKAJIMA YOSHIYUKI,
 PI MAKIGUCHI NOBUYOSHI
 PC C12N15/00, C12N1/20//C12N9/88, (C12N1/20, C12R1/01), (C12N9/88, PC
 C12R1/01);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=IFO 559 (ATCC 10788);
 FH Location/Qualifiers
 FH

FT CDS 1..2151
 FT
 3'UTR 2152..2331.
 FEATURES
 source location/qualifiers
 1..2331 /organism="Rhodospiridium toruloides"
 /mol_type="genomic DNA"
 /db_xref="taxon:5286"
 ORIGIN
 Query Match 67.8%; Score 1677.4; DB 6; Length 2331;
 Best Local Similarity 80.4%; Pred. No. 1.3e-174; Indels 36; Gaps 6;
 Matches 1891; Conservative 0; Mismatches 425;
 1 ATGGCCCCCTCCNTCGATCGGACCTCGTCTGCGCAAGCGGCTGCGGAGGTTG 60
 Db 1 ATGGACCTCGCTCGATCGATCTCGGACCTGTTCCAAACGGCGTGGATCGCAAG 60
 QY 61 CACGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 120
 Db 61 CAGGCTGT-----CAATGGCGCTTGAACCACTC-----GCATGCGAG 100
 QY 121 GCTGCTCTCCCGACACCGAGNNAGAGAGCTGATGATGATGATGATGATGATGATGAT 180
 Db 101 GCTGCGACCTGCGCCACACCGAGGTCAGAGTCAGATGTCAGATGTCAGATGTCAG 160
 QY 181 ACCCGACCGNNACGNNACGNNACGNNACGNNACGNNACGNNACGNNACGNNACG 240
 Db 161 CGCGACCGACCTCGA---CGCTCGAACCTCGAGGCTACTGCTCTAACCTCGAGAC 217
 QY 241 TGGGGCG 300
 Db 218 TCTCGCGCGCGAG 276
 QY 301 AAGATCGACAAAGAGTGAAGTCTCTCCGNNCCGAGCTGNAACAAGAGTCTAGGNGTC 360
 Db 277 AAGATTGAACAATCGTGTGAGTCTTGGCTCGCACTCTCATGAGAGTCTAGGCGTC 336
 QY 361 ACGACTGCTTGGGGGCTCGGCGCGACCCGAGCTAGAGATGCAATCTCGCTCAGAG 420
 Db 337 ACGACTGATTTGGGGGATCCGAGACACCCGACCGAGAGCGCATCTCGCTCAGAG 396
 QY 421 GGNCTCTCGAGACACCACTGCGGTTGCTCCGNNCCGAGCTGNAACAAGAGTCTAGGNGTC 480
 Db 397 GCTCTCTCGAGACACCACTGCGGTTGCTCCGTTGCTGCTGCTGCTGCTGCTGCT 456
 QY 481 GGNCGCGCTCGAGAACTCGTTCCGTCGAGAGTCTGCGCGCGCGCGCATACCATCCG 540
 Db 457 GCGCGGCTCTGAGAACTCGTTCCGTCGAGAGTCTGCGCGCGCGCATACCATCCG 516
 QY 541 GTCACTGCTGACNCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 517 GTCACTGCTGACNCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 576
 QY 601 AACTCTCTCAACCAAGGATCAACCCCATGTCCTCCCTCGCGCGACCATCTCGGCTG 660
 Db 577 AACTCTCTCAACCAAGGATCAACCCCATGTCCTCCCTCGCGCGACCATCTCGGCTG 636
 QY 661 GCGGACTCTCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 Db 637 GCGGACT 696
 QY 721 GTNCACTGNNTCAGAGAGGAGAGAGATCATGTGCGCGCGCGAGCGCATCGGCTG 780
 Db 697 GTNCACTGNNTCAGAGAGGAGAGAGATCATGTGCGCGCGCGAGCGCATCGGCTG 756
 QY 781 TTNGTCTGAGCGCGTCTCTCTGCGCGCGAGAGAGGTTCTGCTGCTGCTGCTGCT 840
 Db 757 TTCACTCTGAGCGCGTCTCTCTGCGCGCGAGAGAGGTTCTGCTGCTGCTGCTGCT 816
 QY 841 GCGGCTCGGCTCGATGCGCGAGCGCGCTGCTGCGAGCGAGACATGCTCTGCTCTC 900
 Db 817 GCGGCTCGGCTCGATGCGCGAGCGCGCTGCTGCGAGCGAGACATGCTCTGCTCTC 876

QY	901	TCGACAGGCGCTCA	CGGCTCMA	CGGTGAGGGCCAT	GGTGGCCAC	CGCGGCTGGTTC	CAAC	960
Db	877	TCGAGTGGCTCA	CGGCGCAT	TAACGCTGCA	AGGCGCAT	TGGTGGCAG	CGCGGCTGGTTC	936
QY	961	CCNTTCTCTCA	CGACGTCA	CGCGCCTCA	CCCCGAC	CGAGATCG	CGCGCA	1021
Db	937	CCCTTCCTTCA	CGACGTCA	CGCGCCTCA	CCCCGAC	CGAGATCG	CGCGCA	996
QY	1021	CGCA	CGCTCC	TCGAGGGC	CAGC	NNGTGTTGGCGTCC	ACACGAGAGGAGGAGTCA	1081
Db	997	CGCA	AGCTCCTCG	AGGAGAC	CGCTTGGTGTCC	ACCATGAGAGGAGGATCG	AAAGTCTAAG	1051
QY	1081	GAGCAG	AGGAGCAT	TTCTCC	CCAGAC	CGGCTAC	CGGCGTCCGCTCAGTGGTCTC	1141
Db	1057	GACGAC	AGGAGCAT	TTCTCC	CCAGAC	CGGCTAC	CGGCGTCCGCTCAGTGGTCTC	1111
QY	1141	GAGCGCGTCC	GTGACGCA	CATGATTCA	CGCGCCAC	CGCGGTC	CTCGGCTCGAGGGCCGACG	1201
Db	1117	GAGCGCGTCC	GTGACGCA	CCTCATTTAC	CGCGCCAC	CGCGGTC	CTCGGCTCGAGGGCCGACG	1171
QY	1201	TCGACG	ACCGCA	CAACCCGCTCAT	TCGACGTG	AGAA	CAAGANGACCA	1261
Db	1177	TCGACG	ACCGCA	CAACCTCTCAT	TCGACGTG	AGAA	CAAGATTGACCA	1231
QY	1261	TTCCAG	AGCGAG	CGCGCTGTGCG	CMAACG	AGATG	AGAGAGAGTGGCGCTGCGCTG	1321
Db	1237	TTCCAG	AGCTCCGCTGTG	CGC	CAAC	CATG	AGAGAGAGTGGCGCTGCGCTG	1291
QY	1321	GGCA	AGCTCA	ACTTCA	CGCAGCTCA	CCGAGATG	CTCAA	1381
Db	1297	GGCA	AGCTCA	ACTTCA	CGCAGCTCA	CCGAGATG	CTCAA	1351
QY	1381	CCNT	CTCGCTCG	CTCGCTCG	AGGAC	CCNTG	CTCGCTCGCTCGCTCG	1441
Db	1357	CCNT	CTCGCTCG	CTCGCTCG	AGGAC	CCNTG	CTCGCTCGCTCGCTCG	1411
QY	1441	GCGCG	CNGCNG	CNTACA	CTTCCG	AGCTCG	AGCA	1501
Db	1417	GCGCG	CNTGCG	CTGAC	CTTCCG	AGCTCG	AGCA	1471
QY	1501	CAG	CGCG	CNAGAT	TGGGCA	CAACGAG	CGCTCA	1561
Db	1477	CAG	CGCG	CNTGAG	TGGGCA	CAACGAG	CGCTCA	1531
QY	1561	AC	NGCCG	AGGCCA	AGCATCTTTCTCTCT	CCTCG	CA	1621
Db	1537	AC	NGCCG	AGGCCA	AGCATCTTTCTCTCT	CCTCG	CA	1591
QY	1621	CAG	CGCGT	CGACCT	CCGCGCG	ATGAGTT	CGAGT	1681
Db	1597	CAG	CGCGT	CGACCT	CCGCGCG	ATGAGTT	CGAGT	1651
QY	1681	NC	NNCG	CTONT	CNAG	CA	CTTTGG	1741
Db	1657	NC	NNCG	CTONT	CNAG	CA	CTTTGG	1711
QY	1741	CT	CAN	GAGCA	AGGATCA	CAAG	CGGTG	1801
Db	1717	CT	CAN	GAGCA	AGGATCA	CAAG	CGGTG	1771
QY	1801	CT	GAG	CGCG	CGTGG	CA	CGCTTT	1861
Db	1777	CT	GAG	CGCG	CGTGG	CA	CGCTTT	1831
QY	1861	NN	GTCT	CTCC	CCNN	NGC	ANNAG	1921
Db	1837	NN	GTCT	CTCC	CCNN	NGC	ANNAG	1891
QY	1921	CG	CG	CGAG	AA	GGC	CA	1981
Db	1891	CG	CG	CGAG	AA	GGC	CA	1851
QY	1981	CG	CG	CGAG	AA	GGC	CA	1941
Db	1851	CG	CG	CGAG	AA	GGC	CA	1811

	LOCUS	E01784	2331 bp	DNA	linear	PAT 29-SEP-1997
	DEFINITION	DNA sequence coding for L-phenylalanine ammonia-lyase.				
	ACCESSION	E01784				
	VERSION	E01784.1	GI:2170037			
	KEYWORDS	JP 1986317087-A/1.				
	SOURCE	unidentified				
	ORGANISM	unclassified.				
	REFERENCE	1 (bases 1 to 2331)				
	AUTHORS	Fukuhara,N., Yoshino,S., Watanabe,M., Nakajima,Y. and Makiguchi,N.				
	TITLE	CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF ADVENTITIOUS GENE PRODUCT USING SAID CONTROL				
JOURNAL	Patent:	JP 1986317087-A 1	26-DEC-1988;			
	MITSUI TOATSU CHEM INC					
COMMENT	OS	Rhodospiridium				
	PN	JP 1986317087-A/1				
	PD	26-DEC-1988				
	PF	18-JUN-1987	JP 1987152358			
	PI	FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI NAKAJIMA YOSHIYUKI,				
	PC	MAKIGUCHI NOBUYOSHI				
	CC	C12N15/00,C12N9/88,(C12N9/88,C12R1.01);				
	CC	strandedness: Double;				
	CC	topology: Linear;				
	CC	hypothetical: No;				
	CC	anti-sense: No;				
	CC	*source: strain=IFO 559 (ATCC 10788);				
	FT	Key	location/Qualifiers			
	FT	CDS	1..2151			
	FT	'UTR	2152..2331.			
FEATURES	source	location/Qualifiers				
		1..2331				
		/organism="unidentified"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:32644"				
ORIGIN	Query Match	67.8%; Score 1677.4; DB 6; Length 2331;				
	Best Local Similarity	80.4%; Pred. No. 1.3e-174;				
	Matches 1891; Conservative	0; Mismatches 425; Indels 36; Gaps 6;				

Db 2126 |ACGCTCCCTCCAGATGCTCGCTTAGACACTCTTCCCACTCTCGATCCCTTCCATACC| 2185
Qy 2221 |NNNCNNNNNNNNCTTTNGNNNTCGNNTCGTGNCNNNNCGANNNTNNNNNNNN| 2280
Db 2186 |TATCCCGCTGCACTTAGACTCGCTTCTTGTGCACTCGGACTCGCATCGCTTCT| 2245
Qy 2281 |TNNNNCTNNCTNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN| 2340
Db 2246 |TCGTTCTTGCTGCTGCTTC-TAGACCGGTGTCGATTACTTACCTCGAGATTGTGAATCAAGC| 2304
Qy 2341 |NGNNNNNNNNNA 2352
Db 2305 |AGTACCATCCA 2316

RESULT 7
E01785
LOCUS E01785 2331 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA sequence coding for L-phenylalanine ammonia-lyase.
ACCESSION E01785
VERSION E01785.1 GI:2170038
KEYWORDS JP 1988317088-A/1.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Bukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
1 (bases 1 to 2331)
Rukuhara, N., Yoshino, S., Watanabe, M., Yamamoto, K., Suzuki, M. and
Nekajima, Y.
REFERENCE
AUTHORS NAKAJIMA, Y.
TITLE CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
JOURNAL ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
PATENT: JP 1988317088-A 1 26-DEC-1988;
MITSUI TOATSU CHEM. INC.
COMMENT
OS Rhodospiridium toruloides
PN JP 1988317088-A/1
PD 26-DEC-1988
PF 18-JUN-1987 JP 1987152359
PI FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
YAMAMOTO KAORU,
SUZUKI MAKI, NAKAJIMA YOSHIYUKI
PC C12N15/00, C12N9/88, (C12N9/88, C12R1.01);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=IFO 559 (ATCC 10788);
FH Key Location/Qualifiers
FT CDS 1..2151
FT 3'UTR 2152..2331.
FEATURES
source Location/Qualifiers
1..2331
/organism="Rhodospiridium toruloides"
/mol_type="genomic DNA"
/db_xref="taxon:5286"

ORIGIN
Query Match 67.8%; Score 1677.4; DB 6; Length 2331;
Best Local Similarity 80.4%; Pred. No. 1.3e-174;
Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;
Qy 1 |ATGGCCCCCTCCTGATGATCGGAGCTCGGAGTGGCAAGCGGAGTGGGAGTGG| 60
Db 1 |ATGGAGCCCTCGCTGACTCGATCTCGGAGTGGTGGCAAGCGGAGTGGGAGTGG| 60
Qy 61 |CAGCGCGNNCGNCGNCGNCGNCGNCGGAGCGGAGCGTCAACNCTGNGCGCGGCG| 120
Db 61 |CAGGCGTGT-----CAATGCGCGCTCGACCAACTC-----GCASTGGAG| 100
Qy 121 |GCTCGTCTCCCGACCAACCGAGNCGAGCGAGCTCGACATGTGAGNAGATCTCGCG| 180
|-----|

Db 101 |GCTCGACTCGCCCAACCCAGGTCACGAGTGTGACATGTGAGAGATGCTCGCCG| 160
Qy 181 |ACCCGACCCGNNACGAGNNNTGAACTGACGAGGTACACCTTCAGGAGCTGCG| 240
Db 161 |CGCGGACCGACTGCA---CGCTGCAACTGACGAGGTACTGTCTCACTCGAGAGCTG| 217
Qy 241 |TGGGGCGCGGCGCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG| 300
Db 218 |TTCGCGCGCGGAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG| 276
Qy 301 |AAGATGCACAANAGAGTGAAGTCTCGAGTCTCGGNNCAGCTCNAACAAGTCTAGG| 360
Db 277 |AGATTGACAAATCGGTGAGATTCTGTGGCTGCGAATCTTCATGAGAGCTTACGG| 336
Qy 361 |ACGACTGTTTGGCGGCTCGGCGGACACCGGACTAGAGATGATCTGCTTCAGAG| 420
Db 337 |ACGACTGATTTGGGCGGATCCGACAGACCCGCGACCGAGAGCGCATCTCGCTCAG| 396
Qy 421 |GCNCTCTCGAGCACACGACTCTGCGGCTGNNCTCCGNNCGTGGNTGANTCGCTC| 480
Db 397 |GCTCTCTCGAGACACGACTCTGCGGCTGTTTCTCTTCTGCTTCTGCTTCTGCT| 456
Qy 481 |GAGCGCGGCTTCGAGAACTCGCTTTCGCTCGAGGTGCTCGGCGCGCATGACATCG| 540
Db 457 |GCGCGCGGCTTCGAGAACTCGCTTTCGCTCGAGGTGCTCGGCGCGCATGACATCG| 516
Qy 541 |GTCACTCGCTACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG| 600
Db 517 |GTCAAGACTTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG| 576
Qy 601 |AACTTCTCAACACAGGATCAACCCATCGTCCCTTCGCGGCGGACATCTCGGCTG| 660
Db 577 |AACTTCTCAACACAGGATCAACCCATCGTCCCTTCGCGGCGGACATCTCGGCTG| 636
Qy 661 |GGCGACTCTCGCCNCTCTCTTAATCATGCGCGCGGCTACCGGTACCCGGAACNN| 720
Db 637 |GGCGACTCTCTCTCTCTCTTAATCATGCGCGCGGCTACCGGTACCCGGAACNN| 696
Qy 721 |GTNACGCTNNNTNACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 780
Db 697 |GTGACGCTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 756
Qy 781 |TTNGGTCTGAGCCCGGTGCTTCTGCGGCGGAGAGGAGGTCTGCTTCAACGAGC| 840
Db 757 |TTCAACCTCGAGCCCGGTGCTTCTGCGGCGGAGAGGAGGTCTGCTTCAACGAGC| 816
Qy 841 |GCGGCTCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 900
Db 817 |GCGGCTCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 876
Qy 901 |TCGAGGCGCTCAAGGCTTNAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 960
Db 877 |TCGAGTGTGCTCAGGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 936
Qy 961 |CNTTCTCCACGACGTCAGCGGCGCTCACCCGACCCAGATCGAGGTGCGGCAACAT| 1020
Db 937 |CCCTTCTTCAAGCGTACGCGGCGCTTCAACCGAGTCAAGTTCGAGGAGGAGGAG| 996
Qy 1021 |CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1080
Db 997 |CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1056
Qy 1081 |GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1140
Db 1057 |GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1116
Qy 1141 |GAGCGGCTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1200
Db 1117 |GAGCGGCTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG| 1176
Qy 1201 |TCGACGACGAGCAACCGGCTCATGACGCTGAGAGCAAGAGCAACGAGCGGCAAC| 1260
Db 1177 |TCGACGACGAGCAACCGGCTCATGACGCTGAGAGCAAGAGCAACGAGCGGCAAC| 1236
|-----|

```
QY 1261 TTCAGAGCGNCGCTGTGCGNNAACAGATGAGAAAGCTCGCCTCGCNCCTTGCCCTGATC 1320
DB 1237 TTCAGAGCTGCGCTGTGGCCAAACATGAGAAAGACTCGCCTCGGCGCTGCCAATC 1296
QY 1321 GGCAAGCTCAACTTACAGAGAGCTCACAGATGCTCAACGCGCGCATGAAACCGGCGCTN 1380
DB 1297 GGAAAGCTCAACTTACAGAGAGCTCACAGATGCTCAACGCGCGCATGAAACCGGCGCTC 1356
QY 1381 CCNTCTGCTGCTGCTGCGGAGAGCCGCTGCTCTCTATCACTGAAAGGCGCTGACATT 1440
DB 1357 CCTCTGCTGCTGCGGCGGAAAGCCCTCGCTCTCTCACTGCAAGAGGCGCTGACATC 1416
QY 1441 GCGCGNCGNCTAACAATTGAGAGCTCGGNCACCTTGGCAAACCGGNNACGACCAAGTC 1500
DB 1417 GCGCGCGCGCTAACCCTGAGATGGGAAACCTCGCAACCTGTGACGACGATGTC 1476
QY 1501 CAGCGCGGAGATGAGGCAACAGGCGCGCTCAACTGCTGCGNCCTCATCTCGGCGCGCG 1560
DB 1477 CAGCGCGGCTGAATGAGGCAACAGGCGCGCTCAACTGCTGCGGCTCATCTCGGCTGTCG 1536
QY 1561 ACNCGCGAGGCAAGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1537 AGACCGAGATCCAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGCGCGCTGACCTCGGCGGAGATGAGATTCGAGTTCAAGAGAGAGTTGACCCGCTNNTC 1680
DB 1597 CAAGCATGACCTTGGCGCGAGTCAAGTTCAAGTTCAAGAGAGAGTTGAGCCGACATC 1656
QY 1681 NCNCGCTCNTNAGAGACACTTTGGCNCNGCCTCGACGCGCNCNNNNNNNNNNNACGAA 1740
DB 1657 GCTCGCTCATGACAGACACTTTGGCTCCGCGCATGACCGGCTGAACTGCGGAGAG 1716
QY 1741 CTCGNGAGCAAGGTCAACAAGNCCTCAACAAGNCCTCGAGAGAGCACTTCTGTAAGAC 1800
DB 1717 CTCGTGGAAGGTGAACAAGAGCTCGGCAAGCGCTCGAGAGAGCACTTCTGTAAGAC 1776
QY 1801 CTCGAGCGCGCGCTGGGACAGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
DB 1777 CTCGTGCGCGCTGGGACAGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
QY 1861 NNCGTCTCGCCNNNNGCCANNAGTCTCGCTCGCNGCGCTCAAGCGCTGAGAGGTGCGCT 1920
DB 1837 TCGTCAGAGTC-----GCTCTCGCTCGCGCGCTCAAGCGCTGAGAGGTGCGCG 1885
QY 1921 CCGCGGAGAGGCGATCTGCTCAAGCGCGCNCNGTCCGAGACNCCTCTCGAGCGCNCCT 1980
DB 1886 CCGCGAGTGGCATCTGCTCAAGCGCGCAAGTCCGAGAGACCTTCTGAGTCCGCGCGCT 1945
QY 1981 CGTCTGCTGCGCGCGCTCNCGTACTCTCTCGCGCGAGCGCGCTCTCTGTAATCTGTTG 2040
DB 1946 CGACCTGCTGCGCGCGCTCTGTAATCTCTGCGCGAGCTCAAGTCTCTACGCGTTG 2005
QY 2041 TCGCGAGAGAGCTCGGCGTCAAGGCGCGCGCGCGAGCTCTTCTCGGCAAGAGAG 2100
DB 2006 TCGCGAGAGAGCTTGGCGTCAAGGCGCGCGCGCGAGAGCTCTTCTCGGCAAGAGAG 2065
QY 2101 TAGCATTCGACCAAGCTCTCCCGCATCTACGAGGCGCTCAAGNCGCGAGCTCAACC 2160
DB 2066 TAGCATTCGAGCTGAAGCTCTCCAGATCTTACGAGGCGCATCAAGTGGGAGAGTCAACA 2125
QY 2161 ACCTCTCTGTCAGAGTCTCGCNTAGNNCNCNNCNACTCGNTNNNNCCNNNCNC 2220
DB 2126 ACCTCTCTCTCAAGATGCTCGCTTACACTCTTCCCACTCGCATCTCTTCATACCC 2185
QY 2221 NNNCCNNNNNNNNCTTNGNNNTGNNCTGNNCNCNNNNCGAGANTNNCCNNNNNNN 2280
DB 2186 TATTCGCGCTGACCTTTAGAGAGCTCTTCTGTGAGAGTGGAACTGCGACTGCTTCTT 2245
QY 2281 TNNNNCTNNCTNNCTCNCNNNNNANNNGTNNNNNNNNNNNNNNNNNNNNNNNN 2340
DB 2246 TCGTCTGTGCTGCTCTCTC-TAGACCGTGTGGGTATTAACCTCGAATTTGTAATGAAGC 2304
```

```
QY 2341 NGTNNNNNNNN 2352
DB 2305 AGTACCATCCA 2316

RESULT 8
RTPALGNA
LOCUS RTPALGNA 2315 bp mRNA linear PLN 23-MAR-1993
DEFINITION R.toruloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION X12702
VERSION X12702.1 GI:288350
KEYWORDS L-phenylalanine ammonia-lyase; PAL gene.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
1 (bases 1 to 2315)

REFERENCE
AUTHORS Unknown.
JOURNAL Unpublished
COMMENT (Patented) EP 0 260 919 A1.
FEATURES
source
1..2315
/organism="Rhodospiridium toruloides"
/mol_type="mRNA"
/strain="IPO 559"
/db_xref="taxon:5286"
/clone="pSW13"
1..2315
/gene="PAL"
1..2151
/gene="PAL"
CDS
1..2151
/product="L-phenylalanine ammonia-lyase"
/protein_id="CA31209.1"
/db_xref="GI:288351"
/db_xref="GOA:P11544"
/db_xref="SWISS-PROT:P11544"
/translation="MAEPLDSISHSPANGVANSAKQAVNGASTNLVAGSHLPTTQVTO
VDIVEKMAAPLDSLELDGYSLNGDVSAARKGRVAVYDSDEISKIDKSEFLR
SOLSWSVGTTGGGASDRTEDAI SLOKXLEHOLCGVLPSPSDFRGLSENL
PLEVVRGAMTIRVNSLIRGSAVAVLVLEALTNLNGITPIVLRGTIASGDLSP
STLAALISGHPDSKVHVHESKEKILYAREKMLFNIPEVYLGKBEGLVNGRAVA
SNATLALDHAMLSLDSLTAMTVEAMVGHGSHFFLHDVTRPHTLTLEKGNIRK
LLEGRFVHHEERKVDDEGILRODRYPLRTSPQWLGSLIHAHVLTEAQ
STTNPPLVDENKTSIHGNGFOAAVANTMEKTLGLAQIGKLNLTQTEMLNANMR
GLPSCLAEDPSLSYHCKGDLIAAAVTSSEGLHANVYTHVCPAEMANAIVNSLAI
SARPTTESNDVLSILIAHLYCVLAOIDIRLIEEPKRGEPALVSLIDHFGSAMTG
SNRPLEYKRNKTLARLBOPTNSYDLVPRHDAFSRAAGVAVVSVLSLSLAANA
WVAALAESALISLTOVRETFMSASTSPALSYLSPFTQILIYAVREBELVKARQDV
FLGKQEVLTIGSNVSKIYEALIKSRINNVLEKMLA"
2315
/gene="PAL"

polyA_site
ORIGIN
Query Match 67.7%; Score 1676.4; DB 8; Length 2315;
Best Local Similarity 80.5%; Pred. No. 1.7e-174;
Matches 1890; Conservative 0; Mismatches 422; Indels 36; Gaps 6;

QY 1 ATGGCCCCCTCNCNTGAGCTGATCGGACCTCGANTCGGCAAGCGCNCNAACGAGTNG 60
DB 1 ATGGCACTCTCCTGATCTGATCTTCGCACTCGTTTCGAAACGGGTGGCATCCGCAAG 60
QY 61 CAGCGCGNCCGNCNNNNGCNCNAAGCGGCGCGACGCTCAACNCTGNGCGCCGCGNG 120
DB 61 CAGGCTGT-----CAATGCGCGCTCGACCAACTTC-----GAGTGGAG 100
QY 121 GCTCGCTCTCCGCGACCAACCGAGNNAGAGAGCTGACATGTTGAGAGATTCCTCGCG 180
DB 101 GCTCGCACTGCGCCCAACCGAGTCAAGAGGTGACATGCTGAGAAAGATGCTCGCG 160
QY 181 ACCCAACCGNNAAGNAGNNNTGAACTGACGGGTACACCTCAACCTCGAGACGTCG 240
DB 161 CGCGACGCACTGCA---CGTCTGAACTCGAGGCTACTCGCTCAACTCGGAGACGTCG 217
```


QY 241 TCAGCGCGCGNCGCAAGAGGCGCGNCGTCCGCGTGCAGACAGNCGACGAGATCCGCGCA 300
DB 218 TCTCGCGCGGAGAGAGGCGAGGCGCTGTCCGCTCAAGACAG -CGAGAGATCCGCTCA 276
QY 301 AAGATCGACAAGAGNGTCGAGTTCTCCGNNCNCAGCTCMAACAAGNGTCTAGCGNGTC 360
DB 277 AAGATTGACAAATCGGTGAGTTCTTGCGCTCGCAACTCTCCATAGGCTCAAGCGCTC 336
QY 361 ACGACTGGTTTCGGCGGCTCGGCGGACACCGCGAGTGAAGAGATCTCCGTCAGAG 420
DB 337 ACGACTGAGATTTCGGCGGATTCGACAGACCGCGACGAGAGGCACTTCGCTCCAGAG 396
QY 421 GCNCTCCTCGAGCAGCAGCTGTGCGGTGTCGCCNACGTCGNTGANTCCTTCGCTC 480
DB 397 GCTCCTCGAGCAGCAGCTGTGCGGTGTCGCCNACGTCGNTGANTCCTTCGCTC 456
QY 481 GAGCGCGGCTCGAGAGACTGCTTCGCTCGAGGTGCTCGCGGCGCATGACATCCG 540
DB 457 GCGCGGCTCGAGAGACTGCTTCGCCNACGTCGNTGANTCCTTCGCTC 516
QY 541 GTCAACTGCTCAACGCGGCACTGCGNGTCCGCTGCTGCTCGAGGCGCTCAC 600
DB 517 GTCAAGAGCTTGACCGCGGCACTGCGGTGCTGCGCTGCTGCTCGAGGCGCTCAC 576
QY 601 AACTTCTCAACACGAGCATCAACCCCATGCTCCCTTCGCGGCACTTCGCGCTG 660
DB 577 AACTTCTCAACACGAGCATCAACCCCATGCTCCCTTCGCGGCACTTCGCGCTG 636
QY 661 GCGACCTCTCCCGNCTCTCMTACATCGCGCGGCACTCAACGCGCAACCGGCAAC 720
DB 637 GCGACCTCTCTCTCTCTCTCTCAATTCGAGCGGCACTCAACGCGCAACCGGCAAC 696
QY 721 GTNCAGGTNNCAGAGGCGCANNGAAGATCATGTCGCCGCGAGCGCATCGCGCTC 780
DB 697 GTGCACTGCTGTCACGAGGCAAGAGAAAGATCTGTAGCGCGGAGGAGCATGCGCTC 756
QY 781 TTMGTCTGAGAGCGCTGCTCTCGCGCGGAGAGGCTTCGCTGCTCAAGGCGAG 840
DB 757 TTTAACTCGAGCGCTGCTCTCTCTCAATTCGAGCGGCACTCAACGCGCAACCGGCAAC 816
QY 841 GCGCTCTCGCGCTCGATGAGCGACCTCGCTGTCAGAGCAGACATCTCTCGCTC 900
DB 817 GCGCTCTCGAGCATGATGAGCGACCTCGCTGTCAGAGCAGACATCTCTCGCTC 876
QY 901 TCGAGAGCGCTCAAGCGCTCTMAAGGTGAGGCGCATGTCGCGCAACGCGGCTGCTCAC 960
DB 877 TCGAGAGCGCTCAAGCGCTCTMAAGGTGAGGCGCATGTCGCGCAACGCGGCTGCTCAC 936
QY 961 CCMTCTCTCAAGCGCTCAAGCGCGCTCAACCGGAGTGAAGTGGCGCGCAATC 1020
DB 937 CCMTCTCTCAAGCGCTCAAGCGCGCTCAACCGGAGTGAAGTGGCGCGCAATC 996
QY 1021 CGGACGCTCTCGAGGCGAGCANNGTTTGGCGTCAACGAGAGGAGTCAAGTCAAG 1080
DB 997 CGGACGCTCTCGAGGCGAGCANNGTTTGGCGTCAACGAGAGGAGTCAAGTCAAG 1056
QY 1081 GAGAGAGAGGAGTTCCTCGCGAGAGCGCTACCGCGCTCGCGAGTGGCTCAATGCTC 1140
DB 1057 GAGAGAGAGGAGTTCCTCGCGAGAGCGCTACCGCGCTCGCGAGTGGCTCAATGCTC 1116
QY 1141 GCGCGCTGCTGAGGAGCATGATTCAGCGCGCAGCGNGTCTCTCGCTCGAGGCGGAG 1200
DB 1117 GCGCGCTGCTGAGGAGCATGATTCAGCGCGCAGCGNGTCTCTCGCTCGAGGCGGAG 1176
QY 1201 TCGAGAGCGAGCAACCGCTCATTCAGAGTTCAGAGCAAGAGACCCACAGCGCGGCAAC 1260
DB 1177 TCGAGAGCGAGCAACCGCTCATTCAGAGTTCAGAGCAAGAGACCCACAGCGCGGCAAC 1236
QY 1261 TTCAGAGCGAGCGCTGTCGNNACAGAGTGAAGAGTCTCGCTCGAGCTCGCGCTGATC 1320
DB 1237 TTCAGAGCGAGCGCTGTCGNNACAGAGTGAAGAGTCTCGCTCGAGCTCGCGCTGATC 1296

QY 1321 GCGAGCTCAACTTCAGCGAGCTACCGAGATGCTCAACGCGCGGATGAACCGCGGCTN 1380
DB 1297 GCGAGCTCAACTTCAGCGAGCTACCGAGATGCTCAACGCGCGGATGAACCGCGGCTN 1356
QY 1381 CCMTCTCTGCTGCTGCGGAGAGCCGCTGCTCTCTATCACTGCAAGGCGCTGACAT 1440
DB 1357 CCMTCTCTGCTGCGGAGAGCCGCTGCTCTCTATCACTGCAAGGCGCTGACAT 1416
QY 1441 GCGCGGAGGAGTCACTTCGAGGTGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1500
DB 1417 GCGCGGAGGAGTCACTTCGAGGTGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1476
QY 1501 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1560
DB 1477 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1536
QY 1561 ACGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1620
DB 1537 ACGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1596
QY 1621 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1680
DB 1597 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1656
QY 1681 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1740
DB 1657 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1716
QY 1741 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1800
DB 1717 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1776
QY 1801 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1860
DB 1777 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1836
QY 1861 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1920
DB 1837 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1885
QY 1921 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1980
DB 1886 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1945
QY 1981 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2040
DB 1946 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2005
QY 2041 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2100
DB 2006 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2065
QY 2101 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2160
DB 2066 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2125
QY 2161 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2220
DB 2126 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2185
QY 2221 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2280
DB 2186 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2245
QY 2281 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2340
DB 2246 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2304
QY 2341 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2400
DB 2305 CAGCGGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 2360

RESULT	9
LOCUS	AR199602
DEFINITION	Sequence from patent US 6355468.
ACCESSION	AR199602
VERSION	AR199602.1
KEYWORDS	GI:20249676
	2419 bp DNA linear
	PAT 20-APR-2002

```

JOURNAL Patent: US 635548-A 12-12-MAR-2002;
FEATURES Location/Qualifiers
SOURCE 1..2419
/organism="unknown"
/mol_type="unassigned DNA"

```

QY	721	GTNCAAGCTNNNTNACAGAGGGCANNAGAAATCATGTNCGCCCGGAGGCGATCGCGCTC	780
QY	721	GTNCAAGCTNNNTNACAGAGGGCANNAGAAATCATGTNCGCCCGGAGGCGATCGCGCTC	780
Db	751	GTTCAAGCTTTTGGACAGAGGAAACCGAAGAAATCATGTTTGGCGGAGAGCCATCTTCGCTC	810
QY	781	TTNAGTCTCGAGCCCGGCTCTCGGACCGGAAGAGAGGCTGTGATCTGTGTAACAGGACAG	840
Db	811	TTTGGTCTCGAGAGGATGCTCTTCGCGCGGAAGAGAGGCTCTGATCTGTGTAACAGGAAAG	870
QY	841	GCCGTCTCCGCTTGATATGAGGACCCCTTCGCTCTGCAAGACACAATGCTCTCGCTCTC	900
QY	901	TGCGAGGCGCTCAAGGCTCTNAGGATCGAGAGGCGCATGTCGAGCCAGGCTTCGCTCAC	960
Db	871	GCCGCTCTCCGCTTGATATGAGGACCCCTTCGCTCTGCAAGACACAATGCTCTCGCTCTC	930
QY	931	TGCGAGGCGCTCAAGGCTCTCAAGGATGAGAGGCAATGTCGCGACAGAGGCTCTTCGCG	990
QY	961	CCNTTCTTCACAGCTGTCACGCGCCCTTCAACCGAACCCAGATCGAGGTCGCGCGCAATC	1020
Db	991	CCGTTTCATCCACAGCTCTGCTCCGCGCGACCCCGGCGAGGTGAGGTGCGCGCGCAATC	1050
QY	1021	CGACAGCTCCCGAGAGGCGACGNNGTTTGGCGTCTCACAGAGAGAGATCAAGGTCAAG	1080
Db	1051	CGACAGCTCTTTCGCGCTGTGCTTTGGCTTTGAGCACAGAGAGAGGTCAAGGTCAAG	1110
QY	1081	GACGACGAGGCAATTCCTCCGACAGACCGGTACCCGCTCCGACGTCGCTCAATGAGCTC	1140
Db	1111	GACGACGAGGCAATTCCTCCGACAGACCGGTACCCGCTCCGACGCTCCGCTCAATGAGCTC	1170
QY	1141	GCGCGGCTCTGACGCGACGATTTGACAGCGCACGCGTCTTCGCTGAGAGCGCGGAG	1200
Db	1171	GCGCGGCTCTGAGGAGCATGATGACGCGCTCTGACTCTTCTGCTGAGAAC--AAC	1227
QY	1201	TCGACGACCGACAACCCGCTCATCGAGCTGACGACAGANAGCCACACGCGCGGCAAC	1260
Db	1228	ACGACGACCGACAACCCGCTCTCTGAGGTGAGAACAGAGGCTCGCACTCGCTCTATC	1287
QY	1261	TTCCAGGCGGNCCTGTGTCGCAACAGATGAGAGAAATCTCGCTCGCNCCTGATC	1320
Db	1288	TTCCAGGCGGTCGTGCTGATTTGATGAGAGAAACAGAGGCTCGCACTCGCTCTATC	1347
QY	1321	GCGAAGCTCAATTCACGAGCTCACCGAGATGCTCAAGCGCGGATGAACCGCGGCTTN	1380
Db	1348	GCGAAGCTCAATTCACGAGCTGACCGAGTGAACCGAGTTGCTCAAGCTGCAATGAACGCGGCTG	1407
QY	1381	CCNTCTGCTCGCTGCTGCGAGAACCCNTGCTCTCTATCACTGAAGGCTCTGACATT	1440
Db	1408	CCTTCTGCTCTCGCTGCTGCGAGAACCCGCTCAATCAACGAGGAGGCTTGACATT	1467
QY	1441	GCGCGGNCGNTTCACTTGAGAGCTCGAGCACTTGCCAAACCGCGTNACAGCAACATC	1500
Db	1468	CACATCGTGTTAAGCTTTGAGAGCTTGCGCACTTGCCAAACCCGCTCAATCACTTGCTC	1527
QY	1501	CAGCGCGNAGATGAGGCAACAGAGCCGTCAACTGCTCGCTGCNCTCATCTCGGANCGCCG	1560
Db	1528	CAGCGCGNAGATGAGGCAACAGAGCCGTCAACTGCTCGCTCATCTCGCGCGCGCGC	1587
QY	1561	ACNCGCGAGGCGAACGAGCTCTTTCTCTCTCTCGCACCCCACTTACGTGCTNCTC	1620
Db	1588	ACTGCGAGGCGCAACGAGCTCTTTCTCTCTCTCGCACCTGATACGACGCTC	1647
QY	1621	CAGGCGCTCGACTCCCGCGAGATGAGATTGAGTTCAAGAACAGTTGACCGCGTNNNTC	1680
Db	1648	CAGGCGCTCGACTCCCGCGAGATGAGATTGAGTTCAAGAACAGTTGACCGCGTCTTC	1707
QY	1681	NCNNGCTCNTCAAGCAGCACTTTGGCNCNGCCCTTGACGGCGNNNNNNNNNNNNAAGAA	1740
Db	1708	CCGATCTCTCCACAGCAGCACTTCGCGACCTGCGCTCGAGT-----AACCA	1755
QY	1741	CTCGAGGACAAGGTCAACAAGNCGCTCNAACAAGCNCCTGACACAGCAACTCTGTAACAC	1800
Db	1756	CTTCGCTCGAGGCTCAAGAAAGGCGCTCAACAGCGCTCTGAGACACACACAGCTGACAC	1815
QY	1801	CTCGAGCCGCGCTGACAGAGCCCTTCTGCTTCGACCCGACCGGACCGCTGTCGAGNNNNN	1860

Db 1816 CTGAGCGCGGCTGGAGACGCGCTTCTGTAAGCGACCGGACCGCTGTCGAGCTCTC 1875
Qy 1861 NNCTCTCGCCNNNNNGCANNAGGTCTGCTCGCNCGCGTCAACGCTGGAAAGGCGCT 1920
Db 1876 TCGTCTCGCCCTCTGCA--AGCTCACCTTACTGCGCTCAACCGCGTGAAGGTGCT 1933
Qy 1921 CCGCGGAAAGGCAATCTGCTTCAACGCGCANGTCGCGACNCCTTTCGNCGCGCCT 1980
Db 1934 CGGCGGAAAGGCAATCTGCTTCAACGCGCANGTCGCGACNCCTTTCGNCGCGCCT 1993
Qy 1961 CGGCGGCGCGCGCGCTCNCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Db 1994 CTTCGAGGCG 2053
Qy 2041 TCCGCGGAGCTCGGCGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
Db 2054 TGGCGGAGGAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2113
Qy 2101 TGAAGATCGGCAACGCTCTCCCGCATCTAGAGGCGCATCAAGNCGCGCATCAAC 2160
Db 2114 AGAGATCGGCGGACCAAGCTCTGCGCATCTAGAGGCGCATCAAGGCGCGCATCAAC 2173
Qy 2161 ACGCTCTGCTCAAGATGCTCGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2220
Db 2174 ACGTCTCTGCTCAAGATGCTCGCTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2232
Qy 2221 NNN 2280
Db 2233 ACCCGAAGACGACGCTTTGCGAGCGTGTGTCGCAAGACGAGCTTCTCTCATACACA 2292
Qy 2281 TNN 2340
Db 2293 TGTGCGCTTACTCTGCGCGCTGATCAGCTCTCTCACTTCTTTCGTAATCCGCGCT 2352
Qy 2341 NGTNN 2400
Db 2353 GGTGCGCA--GTACACGTATATAGAGCTGGAATGATGCAAGTCTTGAAGTCAAAAA 2411
Qy 2401 NNNNNAAA 2408
Db 2412 AAAAAAAA 2419

RESULT 10
AX36860 2419 bp DNA linear PAT 15-FEB-2002
LOCUS AX36860
DEFINITION Sequence 12 from Patent WO0208402.
ACCESSION AX36860
VERSION AX36860.1 GI:18698149
KEYWORDS
SOURCE Rhodotorula graminis
ORGANISM Rhodotorula graminis
Eukaryota; Fungi; Basidiomycota; Uredinomyces;
Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
Rhodotorula.
REFERENCE 1
AUTHORS Yoshida,R.K. and Kootstra,A.B.
TITLE Phenylalanine ammonia lyase polypeptide and polynucleotide
sequences and methods of obtaining and using same
JOURNAL Patent: WO 0208402-A 12 31-JAN-2002;
PCBU Services, Inc. (US)
FEATURES
source 1..2419
Location/Qualifiers
/organism="Rhodotorula graminis"
/mol_type="unassigned DNA"
/db_xref="taxon:29898"
37..2199
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD23828.1"
/db_xref="GI:18698150"
/db_xref="REMBTREMBL:CAD23828"

/translation="MAPSLSLATTTANGTNGSHAAPTSAAQPTSLARTRPGLDGH
AAHQSLEIVELSDPDDIVELSGSLTVRDVVGARGRVAVQNDDELARVDK
SVDPLKQNSVYGVTTGFGSADTREVLSLOKALIEHQLEPTPSXSFSGYR
GLENTLPLEVVGAMVIRVNSLTGHSAGVAVLVEALTNLRNIRPIVLRGSIAS
GDLSPSLYIAGITGHPDVKRVHTEGTEKIMFARBAISLFGLEAVLVGPKGLV
GTAASASATLSLIDSHNLSLSQALTLTALVEAAYGQGSFARPIHDVCPHGOYEV
ARNRTLLSGSSPFAVEHEEKVQDDEGTHRODRIPRTSPQSLGPEVEMHARISL
SLENNITDNPDLVENVKQTAHGNFQASVVISMEKTRIALALIGLKNITCTBLN
AAMRGSPSCIAEDPSLNYHKGJLDHIAVAVSLGHLNLPVTVQAPKEMQAVN
SLAISARTIAENDVLSLASHSLTCLQAVDLRAEMLDFKQFPLPLTLOQHLG
TGLDYNALALEKVALNKRLBQTTLYLEPRMDAPFAYATGVVELSSPSANVTLT
AVNMKVASAEKALSILREVRNFMQWPSQADPAHAYLSFRTRVLVSFVREELGVAR
RGDFVGVQDETIGSNVRIYAIKDRINHVLMKLA"
493..495
/note="Other information: Xaa = Val or Ala"

ORIGIN

Query Match 65.8%; Score 1629; DB 6; Length 2419;
Best Local Similarity 78.0%; Pred. No. 2,58-169;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

Qy 1 ATGGCCCCCTCCCTGACTGATCGGACCTCGATGCGCCAAAGCGCNTCNAAAGNTNG 60
Db 37 ATGGCCCCCTCCCTGACTGATCGGACCTCGGACCAAGCGCTTACCAAGCGCTCG 96
Qy 61 CAGCGCGGCG 120
Db 97 CAGCGCGGCTCGGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154
Qy 121 GCTGCTCTCTCCCGACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 155 GCGTGATAGGCGACG 214
Qy 181 ACCCGACCGGNAAGAGGNNNTTCAACTCGAGGCTTCAACCTTCAAGCTGAGAGTGG 240
Db 215 ACCCGACCGGAC--GACGCTGTGAGGCTCAGCGGGTCAAGCTTCAAGCTGAGAGTGG 271
Qy 241 TGGGCG 300
Db 272 TGGGCG 330
Qy 301 AAGATCGACAAAGAGTGAAGTCTCTCGGNNCAGCTCNACAAAGAGTCTAGCGGNTC 360
Db 331 CGGCTCGACAAAGAGTGAAGTCTCTCGGNNCAGCTCNACAAAGAGTCTAGCGGNTC 390
Qy 361 AAGATGATTTGCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 391 ACCAGGAGTTTGGGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
Qy 421 GCNCTCTCGAGACCAAGCTTGGGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 451 GCGCTCATCGAGACCAAGCTTGGGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510
Qy 481 GCGCGCGCGCTCGAGAACTCGCTTGGGCTGTCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 511 GAGCGCGCGCTCGAGAACTCGCTTGGGCTGTCGCGCGCGCGCGCGCGCGCGCGCG 570
Qy 541 GTCACTGCTCAGCG 600
Db 571 GTCACTGCTCAGCG 630
Qy 601 AACTTCTCAACAGCGAGTCAACCCCATGTCCTTCCGCGCGCGCGCGCGCGCGCGCG 660
Db 631 AACTTCTCAACAGCGAGTCAACCCCATGTCCTTCCGCGCGCGCGCGCGCGCGCGCG 690
Qy 661 GCGGAGCTCTCCGCGCTCTGATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 691 GCGGAGCTCTCCGCGCTCTGATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
Qy 721 GTNACGTTNNTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 751 GTTCACTTTTTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810

QY	781	TTNNGGCTCGAACCCTGTCCTCGGCCCCAAGAGAGGTCTCGTCTCGTCAACGGCACG	840
Db	811	TTTGGTCTCGAGGAGCTGTCTCTGGCCCCAAGAGAGGTCTCGTCTGATCAAGCAACG	870
QY	841	GCCGTCTCCGCTCGATGGGAGCCCTTCTGCTGCAAGACGACACATGCTCTGCTCTC	900
Db	871	GCGGCTCTCGGCTCGATGGCAGCCCTCATGCTGCAAGACTCGACATGCTCTGCTCTC	930
QY	901	TGGCAGAGGCGCTCAGGGCTCTMAAGGTGAGAGCATGTGCGGCCACCGCGGCTCGTTCCAC	960
Db	931	TGGAGGCTTGAACGGCTCTTCAAGGTGAGAGCATGTGCGGCCACCGCGGCTCGTTCCAC	990
QY	961	CCNTTCTTCACAGCAGTTCACGGCCCTCACCCGACCCAGTTCGAGGTTCGCGCAATC	1020
Db	991	CCGTTCTATCAAGAGTTCGCTGCGGCCGACCCCGGCCAGGTTCGAGGTTCGCGCAATC	1050
QY	1021	CCGACGCTCTCGAAGGAGAGCMMTTTGGCCGTCCACACAGAGAGAGGTTCAGAGTCAAG	1080
Db	1051	CCGACGCTCTTTCGGGTCTGCTCTTTCGTTGAGTGAACAGAGAGAGGTTCAGAGTCAAG	1110
QY	1081	GACGACGAGAGGATTTCTCCGACAGAACCGGTACCCGCTCGACAGTCCGCTTCAGTGGTCTC	1140
Db	1111	GACGACGAGAGGATTTCTTCCGACAGAACCGGTACCCGCTCGACAGTCCGCTTCAGTGGTCTC	1170
QY	1141	GCGCGGCTCTGAGGAGATGATTTCAAGCCACGCGTCTCTCGTCTGAGGCGGAGACG	1200
Db	1171	GCGCGGCTCTGAGAGATGATGATCAAGCTTCACTGATCTCTCGTCTCGAAGAC---AAC	1227
QY	1201	TGCGACGACCGAACACCCGCTCATGAGCTGAGAGAACAGAGAACCCACACGCGCGCAAC	1260
Db	1228	ACGACGACCGAACACCCGCTCTGAGCTGAGAGAACAGAGAACCGCGCGCGCAAC	1287
QY	1261	TTCCAGAGGCGCTGTCGTCGNAACGATGAGAGAACTGCGCTCGCTGCTGCTGATC	1320
Db	1288	TTCCAGAGGCTGCGCTGTCGTCGATTTGATGAGAGAACAGAGGCTGCACTGCTCTCATC	1347
QY	1321	GGCAGGCTCAATTACGCAAGCTACCGAGATGTCAAGCGCGCATGACCGCGGCGCTN	1380
Db	1348	GGCAGGCTCAATTACGCAAGCTACCGAGATGTCAAGCTGCAATGACGAGAACCGCGGCTG	1407
QY	1381	CCNTCTGCTCTGCTGCGCGAGAACCCCTGCTCTCTATCATCTGCAAGGAGCTTCACATTT	1440
Db	1408	CCNTCTGCTCTGCTGCGCGAGAACCCGCTGCTCTATCATCATGAGGCAAGGCTTGGACATTT	1467
QY	1441	GCGCGGCGCTCAACTCTTGGAGCTGGGACCTTGTCCAACCCCGGTACACCCACGTC	1500
Db	1468	CACATCTGCTCTTACGCTTGGAACTGGGCACCTTGGCAACCCCGTCACTACCTTGTCTC	1527
QY	1501	CAGCGCGGCGATGGGCAACAGAGCGGTAACTCGCTGCGCTCATCTCGCGGCGCGC	1560
Db	1528	CAGCGCGGAGATGGGGAACAGAGCGGTAACTCGCTGCTCATCTCGCGGCGCGC	1587
QY	1561	ACNCGCGAGGCGAACGAGTCTTTCTCTCTCTCTGCGCACCCACTCTTACTGCGGTCTC	1620
Db	1588	ACTGCGAGGCGAACGAGTCTTTCTCTCTCTCTGCGCACCTGTACTGCACTCTC	1647
QY	1621	CAGGCGGTGACCTCCGCGCGATGAGTTCGAGTTCAAGAGGATTCGACCCGATNTTC	1680
Db	1648	CAGGCGGTGACCTCCGCGCGATGAGTTCGAGTTCAAGAGGATTCGACCCGATNTTC	1707
QY	1681	NCNNGCTCTNAGCAGCACTTTTGGCNCAGCCCTCGACGCGCANNNNNNNNNNAACAA	1740
Db	1708	CCGACTCTCTCTCAACAGCACTCTGGCACTGGCTCGAGCTC-----AACCA	1755
QY	1741	CTCGNAGAACAGTGAACAAGNCGCTNACCAAGGNTCGAGAGACCACTCGTAGACAC	1800
Db	1756	CTTGGCTCTGAAGTGAACAAGCGCTGAACAAGGTTTCGAGAGACGACGACGTAGACAC	1815
QY	1801	CTCGAGCGCGCTGCGACAGACGCTTCTGTTTCGCGACCGGACACGCTCGTCAAGNNNNN	1860
Db	1816	CTCGAGCGCGCTGCGACAGACGCTTCTCGTAGAGCACCGGACACGCTCGTAGAGTCTCTC	1875
QY	1861	NNGCTCTGCGCANNNNCCANNAGGTCTGCTCGGNCGCGCTCGAACGCTCTGAAAGTTCGCT	1920

Db		1876	TCGCTCCTCGCCCTGTGGCA--ACGTACCCCTTACTGCCTGCTCAAGCGCTGAAGTTGCC	1933
OY		1921	CCGCGCGAAGAAGCCATCTCGCTCA CGCGCNANGTCCGCGACNCTTTCTGCNGCGNCCT	1980
Db		1934	CGGCGGAAGAAGCCATCTCGCTCA CGGCGAAGTGCGCAACCCTTTGTGACAGCGCGT	1993
OY		1981	CGTCGTCGTGCCCCGCGCCTCNCGTA CTTCTCGCGCGCACGCGCGTCTGTATCGATTG	2040
Db		1994	CTTGCAGAGCGCCGCGCCACGACNA CTCTCTGCGCGCACGCGCGTCTGTATCGTTG	2053
OY		2041	TCCGCGAGAGACTCGCGCTCAAGG CCGCGCGCGACGTTCTTCTGGCAAGCAGAG	2100
Db		2054	TGCCCGAGGAGCTCGCGGTCAAGG CCGCGCGCGCGCGCGTAGTGTGTGCGGTGAGAGG	2113
OY		2101	TGACGATCGGACCAACGTCCTCCG CATCTTACGAGCGCATCAAGNCGCNGCATCAAC	2160
Db		2114	AGACGATCGGAGGACA GCTCTCGGCA TCTACGAGSCATCAAGACGCCGATCAAC	2173
Db		2174	ACGTCCTCGTCAAAATGCTCGCGGT A-AGGCCCGAGCAAGCTCGCCTAGACGCGCGCTC	2232
OY		2221	NNNCNNNNNNNNCTNTGNNTGNNT GNNTGNCNNNNOCGANNTNNNNNNNNNN	2280
Db		2233	ACCCCAAAGCAGTTTTCAGCGTC GTGTGCGCAAGACGATCTTCTTCATACACA	2292
OY		2281	TNNNCNTNNCTNNCTCN CNNNNANCNNGT CNNTNNNNNNCTNNNGNTNNNNNNCNNNC	2340
Db		2293	TGTGCGCTTACTCTCTGCGCGTAT CA CGTCTCGATTCTTGATCCGCGCTCTCTC	2352
OY		2341	NGTNNNCANNACNCNTNNNNNN NANNCGNANGANTTNANGNNTGCGNCCNNNNNN	2400
Db		2353	GGTGTGTA-GTACACGTGTATAGAG CCTGGAATGATTGCAAGTTCGAGTTCAAAAA	2411
OY		2401	NNNAAAA 2408 	
Db		2412	AAAAAAAA 2419	
RESULT 11				
E01759		2151 bp	RNA linear	PAT 29-SEP-1997
LOCUS				
DEFINITION		DNA encoding L-phenylalanine ammonia-lyase.		
ACCESSION		E01759		
VERSION		E01759.1 GI:2170012		
KEYWORDS		JP 1988291583-A/1.		
SOURCE		Rhodospiridium toruloides		
ORGANISM		Rhodospiridium toruloides Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Microsporidomycetidae; Sporidobolales; Rhodospiridium.		
REFERENCE		1 (bases 1 to 2151)		
AUTHORS		Fukuhara,N., Yoshino,S., Yamamoto,K., Nakajima,Y., Iwata,T., Watanabe,M., Suzuki,M. and Wakisuechi,N.		
TITLE		AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME		
JOURNAL		Parent: JP 1988291583-A 1 29-NOV-1988;		
COMMENT		Mitsui TOATSU CHEM INC		
		OS Rhodospiridium toruloides		
		PN JP 1988291583-A/1		
		PD 29-NOV-1988		
		PF 22-MAY-1987 JP 1987123950		
		PI FUKUHARA NOBUHIRO, YOSHINO SADAO, YAMAMOTO KAORU, PI NAKAJIMA YOSHITYUKI,		
		PI IWATA TOMOYUKI, WATANABE MIDORI, SUZUKI MAKI, PI MAKIGUCHI NOBUYOSHI		
		PC C12N9/88,C12N1/16,C12N1/18,C12N1/20,C12N15/00,C12P13/22, PC (C12P13/22,		
		PC C12R1:19), (C12P13/22,C12R1:865);		
		CC strandedness: Double;		
		CC topology: Linear;		
		CC hypothetical: No;		

CC anti-sense: No;
 CC *source: strain=IFO 559;
 CC *source: clone=PSM101;
 FH Key Location/Qualifiers
 FH CDS 1..2151
 FT /product='L-phenylalanine ammonia-lyase' FT
 mat_peptide 1..2148
 Location/Qualifiers
 1..2151
 /organism="Rhodospiridium toruloides"
 /mol_type="genomic RNA"
 /db_xref="taxon:5286"

ORIGIN

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 56-169;
 Matches 182; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGCCCCCTCCNTGCACTGATTCGGACCTCGTGGCCAAAGGCTCNCAACGAGTNG 60
 DB 1 ATGGCAACCTCGCTCGACTCGATCTGCACTCGTGGCAAAGGGCTCGCATCCGCAAAG 60
 QY 61 CAGCGGNNCCGNNNGNCGNNAAGGGCGGCAAGTCCACNCTGNGCGGCGGNG 120
 DB 61 CAGGCTGT-----CAATGGGCTCGACCAACTC---GCAATCGGAG 100
 QY 121 GCTGCTCTCTCCGACACCCAGANNAGCGCAGCTCGATGTTAGNAGATTCCTCGCG 180
 DB 101 GCTGACCTGCGCCCAACCCAGATCAGCAGCTCGATCGATCGAGAAATGCTCGCG 160
 QY 181 ACCCCACCGNAGNAGNNTGAACTGAGAGGGTACACCTCACTCGGNAAGCTCG 240
 DB 161 CGCGACCACTGGA---CGCTGAACTGAGAGGCTACTCGCTCAACTCGGAGACGTG 217
 QY 241 TCGGCGCCGNCAGAGGGCGCGCNCNCTCGGCTCNAGAGNAGNAGATCCGCGCA 300
 DB 218 TCTCGGCGGAGAGAGGGCAGGCGTTCGCGTCAAGAGACAG--GAGAGATCGCTCA 276
 QY 301 AAGATGCAANAAGTGAATTCTCCGANNCAAGCTCAACAAGNCTTACGAGTTC 360
 DB 277 AAGATTGACAAATCGGTGAGATTCTTGCGCTCGCAACTCTCATGAGGCTCTACGGGCTC 336
 QY 361 ACGACTGTTTGGCGGCTCGGCGCAACCCGGAATGAGATGAGATTCGCTCAGAG 420
 DB 337 ACGACTGATTTGGCGGATTCGCGACACCCGCAACCGAGAGCATCTCGCTCAGAG 396
 QY 421 GCNCTCTTGAGACCAAGCTCTGCGGTGTCTCCNACGTGANTCGTCTTNGGCTC 480
 DB 397 GCTCTCTCGAGACCAAGCTCTGCGGTGTCTCCCTTGTGCTTCACTCGTTCGCGCTC 456
 QY 481 GAGCGGCGCTGAGAACTCGTTCGCTCGAGGCTCGCGGGGCGCATGACATCGCG 540
 DB 457 GAGCGGAGTTCGAGAACTCGCTTCCTCGAGGTTGTTGCGGGGCGCATGACATCGCG 516
 QY 541 GTCACTCGCTCAACGCGGCGCACTCGGNGTTCGCTCGTCTGAGGCGCTCAAC 600
 DB 517 GTCAACAGCTTGACCGCGGCGCACTCGGCTGTCGCTCGTCTGAGGCGCTCAAC 576
 QY 601 AACTTCTCAACCAAGGCACTCAACCCCATCTGTCCTCGCGGCAACATTCGGGCTCG 660
 DB 577 AACTTCTCAACCAAGGCACTCAACCCCATCTGTCCTCGCGGCAACATTCGGGCTCG 636
 QY 661 GGGCACTCTCCGCTCTCTAATCAATCGCGCGGCGCATCAAGGTCACCCGCAANNAG 720
 DB 637 GGGCACTCTCTCTCTCTCTAATCAATCGCGGCGCATCAAGGTCACCCGCAAGAG 696
 QY 721 GTTCAAGTNNTCAGAGGGGCAANNAGAAATCATGTTGCGCGGAGGCGATTCGCTC 780
 DB 697 GTTCAAGTNNTCAGAGGGGCAANNAGAAATCATGTTGCGCGGAGGCGATTCGCTC 756
 QY 781 TTNGTCTCGAGCGGCTGCTCTCGGCGCGAAGAGGAGTCTGCTCTGTCACAGCGAGC 840

DB 757 TTCAACCTCGAGCCCGTGTCTCGGCCCGAAGAGATGTCGATCTGTCACAGGACCC 816
 QY 841 GCGGCTCGGCTCGATGATGAGACCTCGCTCTGAGACGAGACATGCTCGTCTCTC 900
 DB 817 GCGGCTCGATGATGATGAGACCTCGCTCTGAGACGAGACATGCTCTGCTCTC 876
 QY 901 TCGAGGCGCTCAAGGCTCTTACGCTGAGGCGCATGAGTTCGAGACAGCGGCTGTTCCAC 960
 DB 877 TCGAGTCTGCTCAAGGCGCATGAGGCGCATGAGTTCGAGACAGCGGCTGTTCCAC 936
 QY 961 CATTCTCTCAAGAGCTGACGCGGCTTCAACCCGCAACCAATGAGGTCGCGGCAATC 1020
 DB 937 CCTTCTCTCAAGAGCTGACGCGGCTTCAACCCGCAACCAATGAGGTCGCGGCAATC 996
 QY 1021 CGCAGCTCTCTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 DB 997 CGCAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
 QY 1081 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 DB 1057 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
 QY 1141 GAGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 DB 1117 GAGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
 QY 1201 TCGACAGCGCAACACCGGCTCATGACGTGAGAGAAAGAGAACCAACCAACCGGCGCAAC 1260
 DB 1177 TCGACAGCGCAACACCGGCTCATGACGTGAGAGAAAGAGAACCAACCAACCGGCGCAAT 1236
 QY 1261 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 DB 1237 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
 QY 1321 GGCAGAGCTCAACTTCAAGCAGCTCAGAGAGTCTCAACCCGCGCATGAAACCGGCGCTN 1380
 DB 1297 GGCAGAGCTCAACTTCAAGCAGCTCAGAGAGTCTCAACCCGCGCATGAAACCGGCGCTC 1356
 QY 1381 CCNTCTGCTCTGCTGCGGAGAGACCTGCTCTCTATATACGCAAGGCGCTCGCATTT 1440
 DB 1357 CCNTCTGCTCTGCTGCGGAGAGACCTGCTCTCTCTCAACGCAAGGCGCTCGCATTT 1416
 QY 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 DB 1417 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
 QY 1501 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 DB 1477 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
 QY 1561 ACNGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 DB 1537 ACNGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
 QY 1621 CAGGCGGTGACCTCGCGGCGATGAGTGAAGTTCAAGAGAGTTCGACCCGANTTTC 1680
 DB 1597 CAGGCGGTGACCTCGCGGCGATGAGTGAAGTTCAAGAGAGTTCGACCCGANTTTC 1656
 QY 1681 NNNGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 DB 1657 GTTCTGCTCATGACGAGCACTTGTGCTCGGCAATGACGAGCTCGAACTCGCGGAGGAG 1716
 QY 1741 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 DB 1717 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
 QY 1801 CTGAGGCGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 DB 1777 CTGAGGCGGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
 QY 1861 NNCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920

Db 1837 TCGTGACGTC-----GCTCTGCTGCGCCGCGTCAAGCCCTGGAAGGTGCGCG 1885

QY 1921 CCGCCGAGAAAGGCATCTCTGCTCAGCGCGCAGNANGTCCGAGACNCTTTGAGCGGCNCCGT 1960

Db 1886 CCGCCCAAGTCGCGCATCTGCTCAGCCCGCAAGTCCGAGACCTTCTCGTCCGCGGT 1945

QY 1981 CGTCGCTGTCGCGCGGCTCNCGTACTCTGCTGCGCGGACGCGGCTCTGTACTGCTTGG 2040

Db 1946 CGACCTGTCGCGCGGCTCTGTACTCTGCGCGGACGCTCAAGTCTTACCGCTTGG 2005

QY 2041 TCGCGAGAGAGCTCGCGCTCAAGAGCCCGCGCGGACGCTCTTCTCGGCAAGAGAGG 2100

Db 2006 TCGCGAGAGAGCTGCGCTCAAGAGCCCGCGCGGACGCTCTTCTCGGCAAGAGAGG 2065

QY 2101 TGAAGATCGCGACCAAGCTCTCCCGCATCTAGAGAGCATCAAGNCGCGCATCAAC 2160

Db 2066 TGAAGATCGCGCTGAAAGCTCTCAAGATCTAGAGAGCATCAAGTCGCGAGATCAACA 2125

QY 2161 AGCTCTCTGTCAGATGCTCGGNTAG 2186

Db 2126 AGCTCTCTCTCAAGATGCTCGTTAG 2151

RESULT 12

AR205032 2151 bp DNA linear PAT 20-JUN-2002

LOCUS AR205032

DEFINITION Sequence 9 from patent US 6368837.

ACCESSION AR205032

VERSION AR205032.1 GI:21502510

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2151)

AUTHORS Gatenby A.A., Satriasani S., Tang X.-S., Qi W.Wei. and Vannelli T.

TITLE Bioproduction of para-hydroxycinnamic acid

JOURNAL Patent: US 6368837-A 9 09-APR-2002;

FEATURES

source 1..2151

Location/Qualifiers

1..2151 /organism="unknown"

ORIGIN

/mol_type="unassigned DNA"

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;

Best Local Similarity 83.7%; Pred. No. 5e-169;

Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGGCCCTCCCTCGACTCGACTCGGACCTCGGACCTCGGCAACGCGTCCGCAAGCGTNG 60

Db 1 ATGGCACCTCTCGCTCGACTCGACTCGACTCGCTGTTGCAAAAGGCGTGCATCCGCAAG 60

QY 61 CAGCGCCGNNCCGNCNNNGCNCNACGCGCGCACGTCACACNCTCNGCNGCGCCGNCNG 120

Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GCAATGCGAG 100

QY 121 GCTTCCTCTCTCCGACCAACCGAGNNNGACGAGCTGAGATCGTNGAGNATCTTTCGCG 180

Db 121 GCTCGACCTGCGCAACCAAGTCAGAGTGAAGATCGTGAAGATGCTTCGCG 160

QY 101 GCTCGACCTGCGCAACCAAGTCAGAGTGAAGATCGTGAAGATGCTTCGCG 160

Db 181 ACCCGACCGNACGNAAGNNNTGAACTGAGGCTAGACCTCTCCGAGAGCTG 240

QY 161 CGCGACCGACCTGA---CGCTCGAACTGAGCGCTACTCGCTCACTCGGAGACGCTG 217

Db 241 TCGGCGCGCGCGCAAGGCGCGCNGTTCGCGCTCNCAGACAGNCGAGATCCGCGCA 300

QY 218 TCTCGCGCGCGAGAAAGGCGAGGCTCTGCTCGCTCAAGAGACG--CGAGAGATCCGCTCA 276

Db 301 AAGATGCAAAANAGTGAAGTTCCTCGGNNCGAGCTCNAACAAGAGTCTACAGAGT 360

QY 277 AAGATTGAATAATCGATCGAGTTCCTGCGCTCGCAACTCTCATAGAGCGTCTTACGCGCTC 336

Db 361 ACGACTGTTTGGCGGCTCGGCGCAACCGGAGTGAAGATGCAATCTGCTCCAGAG 420

Db 337 ACGACTGATTTTGGCGGATTCGAGACACCCGCAOCCAGAGAGGCAATCTGCTCCAGAG 396

QY 421 GNCCTCTCGAGACCAAGCTCTGCGGTGNCCTCCGACGTCGATGANTCTTCCGCTC 480

Db 397 GCTCTCTCGAGACCAAGCTCTGCGGTGNCCTCCGACGTCGATGANTCTTCCGCTC 456

QY 481 GNCGCGGCTCGAATCTGCTCGGTGNCCTCGAGTGCAGCGGCGCATGACCATCGC 540

Db 457 GCGCGGCTCTGAGAACTCGCTTCCCTCGAGTGTTCGCGCGGCGCATGACCATCGC 516

QY 541 GTCAACTGCTCAACGCGGCACTCGGAGTTCGCGCTGCTGCTCTGAGCGCTCAAC 600

Db 517 GTCAACAGTTGACCCGCGGCACTCGGCTGCTCGGCTGCTGCTCGAGCGCTCAAC 576

QY 601 AACTCTCTCAACAGGATCAACCCCAATGTCCTCCCTCGGCGGCAACATCTCGGCGT 660

Db 577 AACTCTCTCAACAGGATCAACCCCAATGTCCTCCCTCGGCGGCAACATCTCGGCGT 636

QY 661 GCGCACTCTCCCNCTCTCTCATGCTGCGCGCATCAACCGTCAACCGGACNNCAAG 720

Db 637 GCGCACTCTCTCTCTCTCTCTCATGCTGCGCGCATCAACCGGATCAACCGGAC 696

QY 721 GTNCAAGTNTTACAGAGGCGANNAGAAATCATGTCGCGCGCGGAGGATCGGCTC 780

Db 697 GTGCAAGTGTTCAGAGAGGCAAGGAGAAATCTGTAGCGCGCGGAGGATCGGCTC 756

QY 781 TTNGGCTCGAGCCCGCTGCTCTCGGCGGAGAGGCTCGGCTCTGTCATCAAGGACG 840

Db 757 TTCAACTCGAGCCCGTGTCTCTCGGCGGAGAGGCTCTGCTCTCTGTCATCAAGGAC 816

QY 841 GCGCTCTCGCTCGATGAGCACTCTGCTGCAACGACACATGCTCTGCTCTC 900

Db 817 GCGCTCTGAGCATGATGAGCACTCTGCTGCAACGACACATGCTCTCTCTCTC 876

QY 901 TCGAGAGGCTGAGGCTCTTACAGGTCGAGGCGCATGTCGCGGCGGCTGCTCTCAC 960

Db 877 TCGAGTGTCTGAGGCGCATGAGGTCGAGGAGTGTGCGGCGGCGGCTCTCTCTCAC 936

QY 961 CATTCTCTCAAGAGTCAAGCGGCTCAACCGGACCGGACCGAGTGAAGTTCGCGGCAATC 1020

Db 937 CATTCTCTCAAGAGTCAAGCGGCTCAACCGGACCGGACCGAGTGAAGTTCGCGGCAATC 996

QY 1021 CCGAGCTCTCTGAGAGGAGCAGNNTTTCGCTCCACCAAGAGAGAGTGAAGTCAAG 1080

Db 997 CCGAGCTCTCTGAGAGGAGCAGGCTTCTCTCTCAACCAAGAGAGAGTGAAGTCAAG 1056

QY 1081 GACGAGAGGAGTCTCTCGGCAAGAGCGCTACCGGCTCCGCAAGTGAAGTCAAGTGCCTC 1140

Db 1057 GACGAGAGGAGTCTCTCGGCAAGAGCGCTACCGGCTCCGCAAGTGAAGTGAAGTGCCTC 1116

QY 1141 GCGCGCTCTGAGCGACATGATTCAGCGCCACGCGNGTCTCTGCTGAGGCGGACG 1200

Db 1117 GCGCGCTCTGAGCGACATTCAGCGCCACGCGGCTCTCAACATGAGGCGGCGAG 1176

QY 1201 TCGAGAGCGCAACACCGCTATGAGAGTGAAGAAAGANAGACCAACGAGGCGCAAC 1260

Db 1177 TCGAGAGCGCAACCTCTCTATGAGCGTCAAGAAAGATTGGACACAGGCGGCAAT 1236

QY 1261 TTCAAGAGCGCGCTGTGCAACGATGAGAAAGATCTGCTGCGCTGCGCTGATC 1320

Db 1237 TTCAAGAGCGCGCTGTGCAACGATGAGAAAGATCTGCTGCGCTGCGCTGATC 1296

QY 1321 GCGAGGCTCAATTCAGCAAGTCAACGAGATGCTCAACGCGGCAATGAAACGCGGCTC 1380

Db 1297 GCGAGGCTCAATTCAGCAAGTCAACGAGATGCTCAACGCGGCAATGAAACGCGGCTC 1356

QY 1381 CATTCTGCTGCTGCTGCGAGGAGACCTGCTCTCTATCACTGCAAGGCGCTGACAT 1440

Db 1357 CATTCTGCTGCTGCGGCGGAGACCTGCTCTCTCTCACTGCAAGGCGCTGACATC 1416

QY 1441 GCGCGGCGGCTCAACTTGGAGTTCGAGGCTCGGACCTTGGCAACCGGCTGACAGTGC 1500

Db 1417 GCGCGGCGGCTCAACTTGGAGTTCGAGGCTCGGACCTTGGCAACCGGCTGACAGTGC 1476

RESULT 13				
AR282639				
LOCUS	AR282639	2151 bp	DNA	linear
DEFINITION	Sequence 9 from patent US 6521748.			PAT 10-APR-2003
ACCESSION	AR282639			
VERSION	AR282639.1	GI:29719238		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 2151)			
TITLE	Tang,X.-S.			
JOURNAL	Poly(nucleotide encoding a mutant Rhodotorula glutinis tyrosine ammonia lyase polypeptide			
FEATURES	Patent: US 6521748-A 9, 18-FEB-2003; Location/Qualifiers			
source	1..2151			
	/organism="unknown"			
ORIGIN	/mol_type="genomic DNA"			
Query Match	65.7%	Score 1626.4;	DB 6;	Length 2151;
Best Local Similarity	83.7%;	Fred. No. 5e-163;		

[illegible]

QY	108	GA	CGA	CGA	GGG	CA	ATT	CT	CCG	CA	GGA	CGG	TAC	CCG	CGT	TGG	CA	CGT	CGT	CGT	CA	AGG	GGT	1140
Db	1057	GA	CGA	GA	GGG	CA	ATT	CT	CCG	CA	GGA	CGG	TAC	CCG	CGT	TGG	CA	CGT	CT	CT	CA	AGG	GGT	1116
QY	1141	GG	CCG	CT	GT	TAG	CGA	CAT	ATT	CA	CGC	CA	CGN	GTC	CT	CT	GT	CGT	CGA	GG	CGN	CAG	1200	
Db	1117	GG	CCG	CT	GT	TAG	CGA	CCT	ATT	CA	GGC	CA	GGC	GT	CCT	CA	CA	TAG	AGG	CG	CGG	CAG	1176	
QY	1201	TG	GA	CGA	CCG	CA	ACC	CT	AT	TGA	AGT	GGA	GA	CA	GA	MTA	CC	CA	CGG	CGG	CAC	1260		
Db	1177	TG	GA	CGA	CCG	CA	ACC	CT	AT	TGA	AGT	GGA	GA	CA	GA	CTT	CGA	CA	CGG	CGG	CAT	1236		
QY	1261	TT	CA	AGG	CGN	CG	CT	GT	G	CN	AA	CAG	CAT	G	AA	G	A	AG	T	G	A	CT	1320	
Db	1237	TT	CA	AGG	CT	G	CG	CT	GT	G	CC	AA	CA	CA	T	G	A	AG	T	G	A	CT	1296	
QY	1321	GG	AA	AG	CT	CA	ATT	CA	CG	CA	GCT	CA	CC	GA	T	GT	CA	AG	CG	GG	AT	GA	1380	
Db	1297	GG	AA	AG	CT	CA	ATT	CA	CG	CA	GCT	CA	CC	GA	T	GT	CA	AG	CG	GG	AT	GA	1356	
QY	1381	CC	NT	CC	GT	CT	G	CT	G	CG	AG	AG	CC	NT	CG	CT	CC	TA	CT	GA	AG	GG	1440	
Db	1357	CC	CT	CT	GT	CT	G	CG	AG	AG	CC	CT	CG	CT	CT	CT	CA	CT	GA	AG	GG	CT	1416	
QY	1441	GC	CG	CG	CG	CN	T	AC	CT	T	G	GA	CT	G	CA	CT	T	G	CA	CC	CG	T	1500	
Db	1417	GC	CG	CT	CG	CG	CT	AC	CT	T	G	GA	CT	T	G	CA	CT	T	G	CA	CC	CG	1476	
QY	1501	CAG	CGG	CG	CA	AT	TGG	G	A	A	C	A	G	CC	G	T	CA	CT	CG	T	C	N	T	1560
Db	1477	CAG	CGG	CT	G	A	AT	TGG	G	A	A	C	A	G	CC	G	T	CA	CT	CG	T	C	N	1536
QY	1561	AC	NG	CC	GA	GG	CA	CA	G	A	G	T	C	T	T	CT	CT	CT	CT	CG	CA	CC	A	1620
Db	1537	AG	GA	CG	AG	CT	CA	A	G	A	G	T	C	T	T	CT	CT	CT	CT	CG	CA	CC	A	1596
QY	1621	CAG	CG	CT	G	A	CT	CT	CG	CG	A	T	G	A	GT	T	CA	A	A	G	A	G	T	1680
Db	1597	CAG	CG	CT	G	A	CT	CT	CG	CG	A	T	G	A	GT	T	CA	A	A	G	A	G	T	1656
QY	1681	NC	NN	CG	CT	NT	CA	AG	CA	CA	CT	T	T	G	C	N	C	CG	CT	G	A	GG	C	1740
Db	1657	GT	CT	GT	CT	AT	T	GA	CA	CA	CT	T	T	G	C	N	C	CG	CT	G	A	GG	C	1716
QY	1741	CT	CG	NG	GA	CA	AG	T	CA	CA	AG	TC	CT	CA	GA	AG	TC	CG	AG	CA	CA	CT	CG	1800
Db	1717	CT	CG	TC	GA	AA	AG	T	CA	CA	AG	TC	CT	CG	CA	AG	TC	CG	AG	CA	CA	CT	CG	1776
QY	1801	CT	CG	AG	CG	CG	T	T	G	CA	AG	CG	CT	T	CT	CG	CA	CG	CG	CA	CG	CT	CG	1860
Db	1777	CT	CG	TC	CG	CG	T	T	G	CA	AG	CG	CT	T	CT	CG	CA	CG	CG	CA	CG	CT	CG	1836
QY	1861	NN	GT	CT	CG	CN	NN	NG																

QY	2161	ACGTCCTCGTCAAGATGCTCGCNTAG	2186
Db	2126	ACGTCCTCGTCAAGATGCTCGCTTAG	2151

RESULT 14			
AX082544			
LOCUS	2151 bp	DNA	
DEFINITION	Sequence 9 from Patent WO011071.	linear	PAT 28-FEB-2001

KEYWORDS	SOURCE	ORGANISM
synthetic construct		
synthetic construct		
artificial sequences.		

AUTHORS Tang, X.S., Vannelli, T.M., Qi, M.W., Sarislan, S. and Garenby, A.A.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent; WO 011071-A 9 15-FEB-2001;
EDITORIAL E.I. DU PONT DE NEMOURS & COMPANY INCORPORATED (US)

```
FEATURES      Location/Qualifiers
source        1. .2151
```

```

/organism="synthetic construct"
/moi_type="unassigned DNA"
/db_xref="taxon:32630"
/note="mutant from Rhodotorula glutinis"

```

ORIGIN

Query Match	65.7%;	Score 1626.4;	DB 6;	Length 2151;
Best Local Similarity	83.7%;	Pred. No. 5e-169;		
Matches 1829;	Conservative	0;	Mismatches 322;	Indels 35; Gaps 5;

ATGCCCCCTCCNCTGACTCGATCGCGAGCTCGNCTGCCAAGGCGNTCNCNAAAGGNTG

[illegible]

Db 61 CAGGCTGT-----CATGGGCTCGACCACCTC-----GCAGTCGAG 100

121 GCTCGCTCTCTCCCGACCA^{CC}CCAGNN^{GA}CGCAGCTCGACATCGTNGAGNAGATCC^{TG}CCG 180

[illegible]

Db 161 CGCCGACGACTCGA---CGCTGAACTCGACGGCTACTCGCTCAACCTTGGAGACGTG 217

241 TGGGCGCCGCGCAAGGGCGCGCNGTCCGCGTCACAGACAGNCAGCAGAGATCCGGCA 300

DU 218 TCTGCGAGGGAGGCGAAGCCATGTCCTCCTTCAACGAGCACG - CGAC GAGGTATCCGCCTCA 27

Ov 301 AACATCGACAAATACGATCGAGTTCCCTCCGCCNCCCAGCTCAAACAATAACGACTCATCGGCGCTC 360

Db 277 AAGATTGACAAATCGGTGAGTTTGGCGTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

361 ACGACTGTTTCGGCGGCTCGGCCACACCCGGACTGAGGATGCNAITCTCGCTCCAGAAG 420

DJ 33 / ACGAC TGGAT TGGGGA TCCGAG CACC CCAC CCGAG GAGC CAT C TCCT CCGAG AG 33

DJ 431 GCGTCCCTCCGACCA CCA CGTCCTCCCTCCGCTGTCCCTCCCA CCGTGCTCCGATTCCCTCCCTCCGCTCC 480

Db 397 GCTCCTCTCGACACCACTCTGCGGTCTCTCCCTTGGTCGTTGACATGTTCCGCTC 456

481 GGNCGGGCCTCGAGAACTGCGTTCGCGCTGAGAGTCTGCGGGGCCATGACCATCCGC 540

[illegible]

Db 517 GTCACAGCTTGACCCGGGCACTGGGCTTCGGCTCGTCTCGAGGCTCACC 576

601 AACTTCCTAACCACGGCATCACCCCATCGTCCCTCCGGGCAACCATCTGGCGTGG 660


```
QY 2041 TCCGCGAGAGAGCTCGGCTCAAGGCCCGCGCGCGAGAGCTTCTCTCGGCAAGCAGAGG 2100
Db 2006 TCCGCGAGAGAGCTTGGCGTCAAGGCCCGCGCGAGAGAGCTTCTCTCGGCAAGCAGAGG 2065
QY 2101 TGACGATCGGACCAACGCTCTCCGCACTCTACGAGGCCATCAAGNNCGGCGNGCATCAAC 2160
Db 2066 TGACGATCGGCTGAAAGCTCTCCAGATCTACGAGGCCATCAAGTGGGCAAGATCAACA 2125
QY 2161 ACGTCTCTGTCAGATGCTCGCTAG 2186
Db 2126 ACGTCTCTCTCAAGATGCTCGCTAG 2151
```

Search completed: September 11, 2004, 23:12:51
Job time : 15249.4 secs

QY 1861 NNCTCTGCGCCNNNNCCANNAGTCTGCTCGCGCCGCTCAACGCGCTGGAAGTGGCT 1920
 DB 1861 NNCTCTGCGCCNNNNCCANNAGTCTGCTCGCGCCGCTCAACGCGCTGGAAGTGGCT 1920
 QY 1921 CCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 1980
 DB 1921 CCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 1980
 QY 1981 CCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2040
 DB 1981 CCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2040
 QY 2041 TCCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2100
 DB 2041 TCCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2100
 QY 2101 TCCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2160
 DB 2101 TCCGCGAGAGAGCCCACTCTGCTCAACGCGCCGCTCAACGCGCTGGAAGTGGCT 2160
 QY 2161 ACCTCTGCTCAAGATGCTCGCTAGNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2220
 DB 2161 ACCTCTGCTCAAGATGCTCGCTAGNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2220
 QY 2221 NNCCNNNNNNNNCTNTTNNNNNTGANTCTGTCNNNNNNCCNNNNNNNNNNNN 2280
 DB 2221 NNCCNNNNNNNNCTNTTNNNNNTGANTCTGTCNNNNNNCCNNNNNNNNNNNN 2280
 QY 2281 TNNNNCTNTTNNCTCNCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2340
 DB 2281 TNNNNCTNTTNNCTCNCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2340
 QY 2341 NCTNNCCANNNAACNTNNNNNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2400
 DB 2341 NCTNNCCANNNAACNTNNNNNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 2400
 QY 2401 NNNNAAA 2408
 DB 2401 NNNNAAA 2408

RESULT 2
 ID AAD45816 standard; DNA; 2475 BP.
 AC AAD45816;
 XX
 DT 07-AUG-2003 (revised)
 DT 27-DEC-2002 (first entry)
 XX
 DE Phenylalanine ammonia lyase consensus DNA.
 XX
 KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
 KW human cytomegalovirus infection; cytostatic; virucide; ds.
 OS Rhodotorula glutinis
 OS Rhodotorula mucilaginosa.
 OS Rhodospiridium toruloides.
 XX
 PN US2002102712-A1.
 PD 01-AUG-2002.
 PF 24-AUG-2001; 2001US-00939408.
 PR 24-JUL-2000; 2000US-00624693.
 PR 24-JUL-2001; 2001WO-US023270.
 XX
 PA (PCBU-) PCBU SERVICES INC.
 XX
 PI Yoshida RK, Kootstra AB;

XX
 DB MPI; 2002-690616/74.
 PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
 PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
 PT human immunodeficiency virus or human cytomegalovirus infection.
 XX
 XX Claim 5; Page 45-46; 74pp; English.
 CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC They are useful for treating mammals having diseases, disorders or
 CC conditions that would benefit from treatment with PAL proteins such as
 CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
 CC human cytomegalovirus (hCMV) infection. Sequences of the invention are
 CC also used in gene therapy. The present sequence is PAL consensus DNA.
 CC (Updated on 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;
 Query Match 86.0%; Score 2126; DB 6; Length 2475;
 Best Local Similarity 100.0%; Pred. No. 2,6e-268;
 Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCCCCCTTCNTGACTGATCGGACCTTCGTCGCCAACGGCCTCNCNAAAGTNTG 60
 DB 1 ATGGCCCCCTTCNTGACTGATCGGACCTTCGTCGCCAACGGCCTCNCNAAAGTNTG 60
 QY 61 CACGGCGNNCCGNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 120
 DB 61 CACGGCGNNCCGNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCC 120
 QY 121 GCTGCTCTCTCCCGACCAACCCAGNNGACGAGCTGACATCGTNGAGNAGTCTCGCCG 180
 DB 121 GCTGCTCTCTCCCGACCAACCCAGNNGACGAGCTGACATCGTNGAGNAGTCTCGCCG 180
 QY 181 ACCCCACCGNNAACGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGAGCTCG 240
 DB 181 ACCCCACCGNNAACGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGAGCTCG 240
 QY 241 TCGGGCGCGCCGCGAAGGGCGCCGCGCTCCGCTCNCAGACAGNCGACGAGATCCGGGCA 300
 DB 241 TCGGGCGCGCCGCGAAGGGCGCCGCGCTCCGCTCNCAGACAGNCGACGAGATCCGGGCA 300
 QY 301 AAGATCGAACAAAGGTGAGTCTCTCCGNNCNCAGCTCNAACAACAGNGTCTACGAGTTC 360
 DB 301 AAGATCGAACAAAGGTGAGTCTCTCCGNNCNCAGCTCNAACAACAGNGTCTACGAGTTC 360
 QY 361 ACGACTGGTTTCGGCGGCTCGGCGGACCCCGACTAGAGTGCNATCTGCTCCGAGAG 420
 DB 361 ACGACTGGTTTCGGCGGCTCGGCGGACCCCGACTAGAGTGCNATCTGCTCCGAGAG 420
 QY 421 GCGCTCTCGAGCACCAAGCTCTGCGGTGTCCTCCNACGTCGNTGCACTCTCNGGCTTC 480
 DB 421 GCGCTCTCGAGCACCAAGCTCTGCGGTGTCCTCCNACGTCGNTGCACTCTCNGGCTTC 480
 QY 481 GCGCGGCGCTCGAGAACCTGCTTCGCTGAGAGTCTCCGCGGCGCCATGACCATTCGCC 540
 DB 481 GCGCGGCGCTCGAGAACCTGCTTCGCTGAGAGTCTCCGCGGCGCCATGACCATTCGCC 540
 QY 541 GTCAACTGCTCANCNCGGGGCACTCGGCGGCTCGCTCGCTCGCTCGAGGCGCTCAC 600
 DB 541 GTCAACTGCTCANCNCGGGGCACTCGGCGGCTCGCTCGCTCGCTCGAGGCGCTCAC 600
 QY 601 AACTTCTCTCAACCAAGGATACCCCATGTCCTCCGCGGCGCCATCTCGGCGTTCG 660
 DB 601 AACTTCTCTCAACCAAGGATACCCCATGTCCTCCGCGGCGCCATCTCGGCGTTCG 660
 QY 661 GCGCACTCTCCCNCTCTNTATCATGCGCGCGCATCACCGGTACCCGAGCANNCAAG 720
 DB 661 GCGCACTCTCCCNCTCTNTATCATGCGCGCGCATCACCGGTACCCGAGCANNCAAG 720

QY 721 GTTNCAGTNNNTACAGAGGCGAAGAAAGATCATGTNCGCCCGGAGGCGATGCGCTC 780
DB 721 GTTNCAGTNNNTACAGAGGCGAAGAAAGATCATGTNCGCCCGGAGGCGATGCGCTC 780
QY 781 TTNGGTCTGAGAGCGGCTGCTCTCGGCGCGAAGAGGAGTCTCGGTCTGTCAACGAGCAG 840
DB 781 TTNGGTCTGAGAGCGGCTGCTCTCGGCGCGAAGAGGAGTCTCGGTCTGTCAACGAGCAG 840
QY 841 GCGGTCTCGGCTCGATGAGCGACCTCGCTGTGACGACGACACATGCTCTGCTCTC 900
DB 841 GCGGTCTCGGCTCGATGAGCGACCTCGCTGTGACGACGACACATGCTCTGCTCTC 900
QY 901 TCGCAGGCGCTCAAGGCTCTNACGAGTCAAGGCGATGAGTCAAGGCGCTGCTCTC 960
DB 901 TCGCAGGCGCTCAAGGCTCTNACGAGTCAAGGCGATGAGTCAAGGCGCTGCTCTC 960
QY 961 CATTCTCTCAAGGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCA 1020
DB 961 CATTCTCTCAAGGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCAAGGCGCTCA 1020
QY 1021 CGCAGGCTCTCGAGAGGCGACGNNGTTGCGCTTCCACGACGAGAGAGGTCAAGTCAAG 1080
DB 1021 CGCAGGCTCTCGAGAGGCGACGNNGTTGCGCTTCCACGACGAGAGAGGTCAAGTCAAG 1080
QY 1081 GACGAGGAGGAGGATTTCTCGGCGAGGACGCTACCGGCTCGGCGAGTCAAGTCAAG 1140
DB 1081 GACGAGGAGGAGGATTTCTCGGCGAGGACGCTACCGGCTCGGCGAGTCAAGTCAAG 1140
QY 1141 GCGCGGCTCTGACGAGGAGTCAAGGCGGACGCGGCTCTCTGCTGAGGCGGAGCAG 1200
DB 1141 GCGCGGCTCTGACGAGGAGTCAAGGCGGACGCGGCTCTCTGCTGAGGCGGAGCAG 1200
QY 1201 TCGACGACCGAGACACCGGCTCATCGAGTCAAGGCGGAGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1201 TCGACGACCGAGACACCGGCTCATCGAGTCAAGGCGGAGAGGAGGAGGAGGAGGAGGAGG 1260
QY 1261 TTTCAGGCGGCGCTGTGCGACGACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
DB 1261 TTTCAGGCGGCGCTGTGCGACGACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
QY 1321 GGGAGGCTCACTTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 1380
DB 1321 GGGAGGCTCACTTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 1380
QY 1381 CATTCTCTGCTCGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
DB 1381 CATTCTCTGCTCGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
QY 1441 GCGCGGCGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1441 GCGCGGCGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
QY 1501 CAGCGGCGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1501 CAGCGGCGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 1561 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1561 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
QY 1621 CAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 1621 CAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
QY 1681 NCNNGGCTCTNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
DB 1681 NCNNGGCTCTNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
QY 1741 CTGAGG 1800
DB 1741 CTGAGG 1800
QY 1801 CTGAGG 1860

DB 1801 CTGAGG 1860
QY 1861 NNGTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
DB 1861 NNGTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
QY 1921 CCGCGGAGG 1980
DB 1921 CCGCGGAGG 1980
QY 1981 CGTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
DB 1981 CGTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
QY 2041 TCGCGAGG 2100
DB 2041 TCGCGAGG 2100
QY 2101 TGAAGATGAGG 2160
DB 2101 TGAAGATGAGG 2160
QY 2161 AGGTCCTGTCAGAGG 2220
DB 2161 AGGTCCTGTCAGAGG 2220
QY 2221 NNCCGAGG 2280
DB 2221 NNCCGAGG 2280
QY 2281 TNNCCGAGG 2340
DB 2281 TNNCCGAGG 2340
QY 2341 NGTNNCCGAGG 2400
DB 2341 NGTNNCCGAGG 2400
QY 2401 NNNNAAAA 2408
DB 2401 NNNNAAAA 2408

RESULT 3
AAD45824
ID AAD45824 standard; DNA; 2163 BP.
XX
AC AAD45824;
XX
DT 27-DEC-2002 (first entry)
XX
DE Yeast phenylalanine ammonia lyase DNA #2.
XX
KW Yeast; phenylalanine ammonia lyase; PAl; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
KW human cytomegalovirus infection; cytoskeletal; virucide; ds.
XX
XX Rhodotorula graminis.
FH
FH Key Location/Qualifiers
FT 1..2163
FT CDS
FT /tag=a
FT /product="PAL protein"
FT /EC_number="4.3.1.5"
FT /transl_except=(pos:13..15, aa:Xaa)
FT /note="Xaa corresponds to Val, Leu, Phe"
FT /transl_except=(pos:34..36, aa:Xaa)
FT /note="Xaa corresponds to Val, Leu, Phe"
FT /transl_except=(pos:46..48, aa:Xaa)
FT /note="Xaa corresponds to Val, Leu, Phe"
FT /transl_except=(pos:49..51, aa:Xaa)
FT /note="Xaa corresponds to Thr, Ala, Ser"

FT /transl_except= (pos:55. .57, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:58. .60, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Leu or an in-frame stop
codon"
FT /transl_except= (pos:73. .75, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:79. .81, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:82. .84, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro"
FT /transl_except= (pos:100. .102, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:106. .108, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:115. .117, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:118. .120, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Gly, Trp"
FT /transl_except= (pos:142. .144, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val,
Gln, Pro, Leu"
FT /transl_except= (pos:160. .162, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:166. .168, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:193. .195, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Val"
FT /transl_except= (pos:196. .198, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:226. .228, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:259. .261, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:277. .279, aa:Xaa)
FT /note= "Xaa corresponds to Asp, Ala"
FT /transl_except= (pos:304. .306, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn"
FT /transl_except= (pos:307. .309, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:325. .327, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:334. .336, aa:Xaa)
FT /note= "Xaa corresponds to Asp, His, Tyr"
FT /transl_except= (pos:340. .342, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:349. .351, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:442. .444, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:448. .450, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:457. .459, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Phe"
FT /transl_except= (pos:460. .462, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp"
FT /transl_except= (pos:469. .471, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Gly, Arg"
FT /transl_except= (pos:475. .477, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:547. .549, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:667. .669, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:673. .675, aa:Xaa)
FT /note= "Xaa corresponds to Ser"
FT /transl_except= (pos:709. .711, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val,
Cys, Phe"
FT /transl_except= (pos:715. .717, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:721. .723, aa:Xaa)
FT /note= "Xaa corresponds to Val"

FT /transl_except= (pos:724. .726, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:735. .738, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn, Thr"
FT /transl_except= (pos:751. .753, aa:Xaa)
FT /note= "Xaa corresponds to Tyr, Ser, Phe"
FT /transl_except= (pos:775. .777, aa:Xaa)
FT /note= "Xaa corresponds to Leu, Phe"
FT /transl_except= (pos:913. .915, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:955. .957, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1036. .1038, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp,
Ser or a stop codon"
FT /transl_except= (pos:1231. .1233, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met"
FT /transl_except= (pos:1261. .1263, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1369. .1371, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:1372. .1374, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1396. .1398, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1459. .1461, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:1477. .1479, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1498. .1500, aa:Arg)
FT /transl_except= (pos:1552. .1554, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:1606. .1608, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1666. .1668, aa:Xaa)
FT /note= "Xaa corresponds to Met, Ile, Val,
Leu"
FT /transl_except= (pos:1669. .1671, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:1672. .1674, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Pro"
FT /transl_except= (pos:1675. .1677, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1681. .1683, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Leu, Phe"
FT /transl_except= (pos:1684. .1686, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:1699. .1701, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1714. .1716, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Tyr"
FT /transl_except= (pos:1723. .1725, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Ala, Val"
FT /transl_except= (pos:1741. .1743, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1747. .1749, aa:Xaa)
FT /note= "Xaa corresponds to Asp, Asp, Tyr"
FT /transl_except= (pos:1849. .1851, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1852. .1854, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:1927. .1929, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
stop codon"
FT /transl_except= (pos:1939. .1941, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:1948. .1950, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:1978. .1980, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:2122. .2124, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
Tyr, Cys"
FT /transl_except= (pos:2128. .2130, aa:Xaa)

/note= "Xaa corresponds to Ser, Arg, Cys"

FT US2002102712-A1.
 XX 01-AUG-2002.
 PD 24-AUG-2001; 2001US-00939408.
 XX 24-JUL-2000; 2000US-00624693.
 PR 24-JUL-2001; 2001MO-US023270.
 XX (PCBU-) PCBU SERVICES INC.
 XX Yoshida RK, Koetsura AB;
 XX WPI; 2002-690616/74.
 DR P-PSDB; AAE27943.
 XX
 PS Claim 7; Page 53-56; 74pp; English.
 CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC
 Query Match 78.5%; Score 1944; DB 6; Length 2163;
 Best local Similarity 93.2%; Pred. No. 1.2e-262;
 Matches 2037; Conservative 0; Mismatches 126; Indels 23; Gaps 7;
 QY 1 ATGGCCCCCTCCNTGACTGATCGCGACCTCGANTGCCAAGCGGCTGNCNAACGANTG 60
 DB 1 ATGGCCCCCTCCBTGCACTGATCGCGACCTCGANTGCCAAGCGGCTGNCNAACGANTG 60
 QY 61 CAGCGCGNCCGNNNGCNGCNAAGGCGCGCAGCTCCAGCTCNGCGCGCGCGCG 120
 DB 61 CAGCGCGCTCCGHCNAAGTCCGSCMAAGGCGCGCAGCTCCAGCTCNGCGCGCGCG 120
 QY 121 GCTGCTCTCCGACCAACCCCAAGGAGAGCTGAGATCTGAGAGATCTCTGCGG 180
 DB 120 GCTGCTCTCTCCCG-CCACCAGVHAGAGAGAGCTGAGATCTGAGAGATCTCTGCGG 178
 QY 181 ACCCCACCGNAGNAGNNTGAACTGAGAGGTTACACCTCACTCGGAGAGCTG 240
 DB 179 ACCCCACCGAGC--GACGWSVTGAACTGAGGAGTTACACCTCACTCGGAGAGCTG 235
 QY 241 TCGGCGCGCNCAGAGGCGCGCNGCTCCGCTGNCAGAGAGAGAGAGAGAGAG 300
 DB 236 TCGGCGCGCGCGCAGAGGCGCGCHBGTCCGCTC-CAGACAGAGAGAGATCCGCGCA 294
 QY 301 AAGATGACAAAGAGTGAATTCCTCCGAGNCCAGCTCCAGAGAGAGAGAGAG 360
 DB 295 AAGATGACAAAGAGTGAATTCCTCCGAGNCCAGCTCCAGAGAGAGAGAGAG 354
 QY 361 ACGACTGTTTGGCGGCTCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 355 ACGACTGTTTGGCGGCTCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
 QY 421 GNCCTCTGAGAGACCAAGCTTGCGGTGTCCTCCNAGCTGAGTGAATCTTCCGCTC 480
 DB 415 GCBCTCTGAGAGACCAAGCTTGCGGTGTCCTCCNAGCTGAGTGAATCTTCCGCTC 474
 QY 481 GAGGCGGCTGAG 540
 DB 475 GAGGCGGCTGAG 534
 QY 541 GTCAGCTGCTCAAG 600
 DB 535 GTCAGCTGCTCAAG 594
 QY 601 AACTTCTCAACAG 660

DB 595 AACTTCTCAACAG 654
 QY 661 GCGGAGCTCTCCGCTCTGNTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 DB 655 GCGGAGCTCTCCGCTCTGNTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714
 QY 721 GTNCAAGTNTTCAAG 780
 DB 715 GTBACAGTNTTCAAG 774
 QY 781 TTNGTCTGAG 840
 DB 775 TTGAGTCTGAG 834
 QY 841 GCGGCTCTCCGCTCTGNTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 DB 835 GCGGCTCTCCGCTCTGNTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
 QY 901 TCGAG 960
 DB 895 TCGAG 954
 QY 961 CCNTTCTCTGAG 1020
 DB 955 CCNTTCTCTGAG 1014
 QY 1021 CGCAGCTCTCTGAG 1080
 DB 1015 CGCAGCTCTCTGAG 1074
 QY 1081 GACGAG 1140
 DB 1075 GACGAG 1134
 QY 1141 GCGGCTCTCTGAG 1200
 DB 1135 GCGGCTCTCTGAG 1191
 QY 1201 TCGAG 1260
 DB 1192 TCGAG 1251
 QY 1261 TTCAG 1320
 DB 1252 TTCAG 1311
 QY 1321 GCGAG 1380
 DB 1312 GCGAG 1371
 QY 1381 CCNTTCTCTGAG 1440
 DB 1372 CCNTTCTCTGAG 1431
 QY 1441 GCGGAG 1500
 DB 1432 GCGGAG 1491
 QY 1501 GCGGAG 1560
 DB 1492 GCGGAG 1551
 QY 1561 GCGGAG 1620
 DB 1553 GCGGAG 1611
 QY 1621 GCGGAG 1680
 DB 1612 GCGGAG 1671
 QY 1681 NCNNCGCTCAG 1740


```

Db      1672 VCBDCGCTGTCVAGCAGCACTTTG3CDCYGCCTCGACGGC-----WAGGAA 1719
QY      1741 CTGGNGGACAAGGTCAACAGAGCGCTCNAACAGCGNCTCGAGCGACCACTCGTAAAGAC 1800
Db      1720 CTGGAGGACAAGGTCAACAGAGCGCTCNAACAGCGNCTCGAGCGACCACTCGTAAAGAC 1779
QY      1801 CTGGAGCGCGCTGGGACGACGCGCTTCTGTTTCGACCGGACCGCTGTCGAGNNNNN 1860
Db      1780 CTGGAGCGCGCTGGGACGACGCGCTTCTGTTTCGAGACCGGACCGCTGTCGAGCTCCTC 1839
QY      1861 NNGTCTCTCCGNNNNGCCANNAGTGTCTGCTGCGNCGCGTCAAGCGCTGGAAGTCCCT 1920
Db      1840 TGTCTCTCCGCTGCTGCA--AGGTCCTGCTGCGCGGTCAAGCGCTGGAAGTCCCT 1897
QY      1921 CCGCCGAGAGGCGCATCTCTGCTCAAGCGGANNCTCGGAGNCTTTCGAGCGGCGT 1980
Db      1898 CCGCCGAGAGGCGCATCTCTGCTCAAGCGGANNCTCGGAGNCTTTCGAGCGGCGT 1957
QY      1981 CGTCTGCTGCGCGCGCTCNCGTAACCTCTGCGCGCGGACCGCGGCTCTGACTGTTCC 2040
Db      1958 CGTCTGCTGCGCGCGCTCNCGTAACCTCTGCGCGCGGACCGCGGCTCTGACTGTTCC 2017
QY      2041 TCCGCGAGAGGCGCTGCGGCGCTCAAGGCGCGCGCGGAGAGTCTTCTCGGCAAGAGAGG 2100
Db      2018 TCCGCGAGAGGCGCTGCGGCGCTCAAGGCGCGCGCGGAGAGTCTTCTCGGCAAGAGAGG 2077
QY      2101 TGAAGATCGGACCAACGCTCTCCGATCTACGAGCGCATCAAGNCGGCGNCGATCAACC 2160
Db      2078 TACAGATCGGACCAACGCTCTCCGATCTACGAGCGCATCAAGNCGGCGNCGATCAACC 2137
QY      2161 ACGTCTCTCTCAAGATGCTCGGCTNG 2186
Db      2138 ACGTCTCTCTCAAGATGCTCGGCTNG 2163

RESULT 4
ABA95243
ID      ABA95243 standard; cDNA; 2439 BP.
AC      ABA95243;
XX      07-AUG-2003 (revised)
DT      10-JUN-2002 (first entry)
XX      XX
DE      R. toruloides PAL polypeptide encoding cDNA.
XX      PAL; yeast; phenylalanine ammonia lyase; cinamic acid; cinamate; ss;
XX      cytotaxic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
OS      Rhodospiridium toruloides.
XX      XX
FH      Key Location/Qualifiers
FT      CDS 1..2151
FT      CDS /*tag= a
FT      CDS /product= "phenylalanine ammonia lyase"
XX      XX
PN      MO200208402-A2.
XX      XX
PD      31-JAN-2002.
XX      XX
PF      24-JUL-2001; 2001WO-US023270.
XX      XX
PR      24-JUL-2000; 2000US-00624693.
XX      XX
PA      (PCBU-) PCBU SERVICES INC.
XX      XX
PI      Yoshida RK, Koester AB;
XX      XX
DR      MPI: 2002-268973/31.
XX      XX
DR      P-PSDB; ABB07693.
XX      XX
PT      Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT      treating mammal having disease or disorder from phenylketonuria, cancer,

```

```

PT      human immunodeficiency virus infection and human cytomegalovirus
PT      infection.
XX      XX
XX      Example 3; Fig 2; 135bp; English.
XX      XX
CC      The invention relates to an isolated and purified yeast phenylalanine
CC      ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC      host cell comprising the polynucleotide is useful for the production of L
CC      -PAL, by adding the polynucleotide or the host cell to a composition
CC      comprising trans-cinamic acid, or trans-cinamate and ammonia, and for
CC      improving the production of PAL, its analog or another optically active
CC      unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC      are useful for treating a mammal having a disease, disorder or condition
CC      selected from phenylketonuria, cancer, human immunodeficiency virus
CC      infection and human cytomegalovirus infection. The present sequence
CC      represents a R. toruloides PAL polypeptide encoding cDNA. (Updated on 07-
CC      AUG-2003 to correct OS field.)
XX      XX
SQ      Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;
Query Match      68.7%; Score 1699.4; DB 6; Length 2439;
Best Local Similarity 79.4%; Pred. No. 1.4e-228;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;
QY      1 ATGGCCCCCTCCNTGACTGATCGGACCTGNTGCGCAACGGGCGTCCNCAACGNTNG 60
Db      1 ATGGACCTCGCTCGACTGATCTCGACATCTTCGCAACAGCGGCTCGCAACG 60
QY      61 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db      61 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 100
QY      121 GCTCGCTCTCTCCGACGACCCAGANNAGCGAGCTGCACTGTTGAGNAGATCTCTCGCG 180
Db      101 GCTCGACCTCGGCGGACCAACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 160
QY      181 ACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db      161 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 217
QY      241 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
Db      218 TCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
QY      301 AAGATCGACCAANAGNGTGAATCTCTCCGANNCGAGTCCNCAACAGNGTCTACGNGTC 360
Db      277 AAGATTGACCAATGAGTGTGAGTCTTGGGCTCGGCACTCTTCATGAGCGTCAAGCGGTC 336
QY      361 ACGACTGTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db      337 ACGACTGATTTGGGCGATCCGACAGACCCGCGACGAGAGCGCATCTCGCTCCAGAG 396
QY      421 GGNCTCTCGAGACCAAGCTCTGCGGTGTCGCGNAGCTCGNAGTCCGANTCGCTTCGCGCTC 480
Db      397 GGTCTCTCGAGACCAAGCTCTGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGT 456
QY      481 GAGCGGCGGCTCGAGACTCGCTTCCGTCGAGGTGTCGCGGCGGCGGCGGCGGCGGCGG 540
Db      457 GCGCGGCGGCTCGAGACTCGCTTCCGTCGAGGTGTCGCGGCGGCGGCGGCGGCGGCGG 516
QY      541 GTCACCTGCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db      517 GTCAACACTTGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
QY      601 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db      577 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
QY      661 GCGGCGGCTCTCCGCTCTCTCATATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db      637 GCGGCGGCTCTCTCTCTCTCATATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 696
QY      721 GTCGCGGTTNNTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780

```

Db 697 GTGACGTGTGTCCACGAGGCAAGAGAAATCTGTACGCCCGGAGGCGATGCGCTC 756
 Qy 781 TTNGTCTCGAGCCCGCTGCTCTCGCCCGAAGAGGCTCTGCTCTGTCAACGAGCAG 840
 Db 757 TTGAACCTGAGAGCCGCTGCTCTCGCCCGAAGAGGCTCTGCTCTGTCAACGAGCAG 816
 Qy 841 GCGCTCTCGGCTCTGATGCGCAACCTTCTCTGTCAACGAGCAGCAGATCTCTCTCTC 900
 Db 817 GCGCTCTGAGCATGATGCGCAACCTTCTCTGTCAACGAGCAGCAGATCTCTCTCTC 876
 Qy 901 TCGCAGAGCGCTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 Db 877 TCGCAGTGTGTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
 Qy 961 CCNTTCTCTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 Db 937 CCTTCTCTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
 Qy 1021 CGACGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 Db 997 CGAAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
 Qy 1081 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 Db 1057 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
 Qy 1141 GCGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 Db 1117 GCGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
 Qy 1201 TCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 Db 1177 TCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
 Qy 1261 TTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 Db 1237 TTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
 Qy 1321 GCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 Db 1297 GCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
 Qy 1381 CCNTTCTCTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 Db 1357 CCTTCTCTCAAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
 Qy 1441 GCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 Db 1417 GCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
 Qy 1501 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 Db 1477 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
 Qy 1561 ACNGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 Db 1537 ACNGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
 Qy 1621 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 Db 1597 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
 Qy 1681 NCNNGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 Db 1657 GTCTGCTCAATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
 Qy 1741 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 Db 1717 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
 Qy 1801 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860

Db 1777 CTGCTCCGCGCTGAGCAGACGCTTCTCTCTGCGCGCGGAGCAGGCTGTGAGTCTCTC 1836
 Qy 1861 NNCTCTCTGCGNNNNNCCANNAGGCTCTGCTCTGCGGCGCTTCAACGCTTGAAGTGGCTT 1920
 Db 1837 TCGTGAAGTC-----GCTCTGCTCTGCGCGCGGAGCAGGCTTGAAGTGGCGG 1885
 Qy 1921 CCGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
 Db 1886 CCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1945
 Qy 1981 CCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
 Db 1946 CCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2005
 Qy 2041 TCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
 Db 2006 TCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2065
 Qy 2101 TCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
 Db 2066 TCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2125
 Qy 2161 AGCTCTCTCAAGATGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
 Db 2126 AGCTCTCTCAAGATGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2185
 Qy 2221 NNCCNN 2280
 Db 2186 TATCCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2245
 Qy 2281 TNN 2340
 Db 2246 TCGTCTTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2304
 Qy 2341 NCTNN 2400
 Db 2305 AGTACCCTCAAGATGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2364
 Qy 2401 NNNNNAAA 2408
 Db 2365 CGCATAAA 2372

RESULT 5
 AAD45815
 ID AAD45815 standard; DNA; 2439 BP.
 XX
 AC AAD45815;
 XX
 DT 07-AUG-2003 (revised)
 DT 27-DEC-2002 (first entry)
 XX
 DE Rhodotorula toruloides PAL DNA.
 XX
 KM Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
 KM human cytomegalovirus infection; cytostatic; virucide; gene; ds.
 OS Rhodospiridium toruloides.
 XX
 FH Key Location/Qualifiers
 FT 1. 2151
 FT CDS /tag= a
 FT /product= "PAL protein"
 FT /EC_number= "4.3.1.5"
 PN
 US20021027.2-A1.
 XX
 PD 01-AUG-2002.
 XX
 PF 24-AUG-2001; 2001US-00939408.
 XX
 PR 24-JUL-2000; 2000US-00624693.

PR 24-JUL-2001; 2001MO-US023270.
XX (PCBU-) PCBU SERVICES INC.
XX Yoshida RK, Koestera AB;
XX MPI; 2002-690616/74.
XX DR P-PSDB; AAE27941.
XX Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.
XX
XX Example 3; Page 41-43; 74pp; English.
XX The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.
CC They are useful for treating mammals having diseases, disorders or
CC conditions that would benefit from treatment with PAL proteins such as
CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC human cytomegalovirus (HCMV) infection. Sequences of the invention are
CC also used in gene therapy. The present sequence is R. toruloides PAL DNA.
CC (Updated on 07-Aug-2003 to correct OS field.)
XX
SQ Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;
Query Match 68.7%; Score 1699.4; DB 6; Length 2439;
Best Local Similarity 79.4%; Pred. No. 1.4e-228;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;
1 ATGGCCCCCTCCNTGACTCGATCGGACCTCGNTGCGCAAGCGGANTCGMAACGGATNG 60
1 ATGGGACCTCGCTGATCGATCTCGACCTCGTTGCGMAAGCGGCTGCGATCGGCAAG 60
61 CACGCGGANNCCNNGNNGCNAAGCGGCGGACGCTCACTGACNTGNGCGCGCGNG 120
61 CAGGCTGT-----CATGGCGCTCGACCACTTC-----GAGTCCGAG 100
121 GGTGCTCTCTCCGACGACCCGAGGAGCGAGCTGACATGTAAGAGATCTCTGCGCG 180
101 GCTGCGACCTGCGCCCAACCCAGGTCAGCGAGTGCATGTCGAGAGAGATGCTGCGCG 160
181 ACCCGACCGANNAGCAGANNNTGAACTCGAGGAGTACCCCTCGGNGACGTCG 240
161 CGCCACGACGACTCGA---CGCTCGAAGTCAAGGCTACTGCTCAACTCGGAGAGCTG 217
241 TCGGCGCGCGGCGGAGGCGGCGGCGGCTGCGGCTGCGAGAGCGAGATCGGCGCA 300
218 TCTCGGCGCGGAGAGGCGGCGGCTGCGGCTGCGAGAGAG---CGAGGAGATCGGCTCA 276
301 AAGATCGACAANAGTGAAGTCTCTCGANNCGAGCTCNAACAAGAGTCTAGCGNGTC 360
277 AAGATTGACAATCGTGTGAGTCTCTGCGGCTGCGAATCTCTCGAGACGCTCTAGCGGCTC 336
361 ACGACTGCTTGGCGGCTCGGCGGACACCGGAGTGAAGTGCATCTGCTTCAGAG 420
337 ACGACTGATTTGGGAGATCGGAGACACCGGACCGAGGAGCGCATCTGCTCCAGAG 396
421 GGNCTCTCGAGACGACGACTCTGGGCTGTCCTCCGACGTCGANTCGANTCTTCGNGCTC 480
397 GCTCTCTCGAGACGACGACTCTGGGCTGTCCTCTCTGCTGCTTCGACTGCTTCGCTC 456
481 GAGCGGCGCTCGAGACTGCTGCTTCGCTCGAGTGTCTCGGCGGCGCATGACATCTCGC 540
457 GAGCGGCGCTCGAGACTGCTGCTTCGCTCGAGTGTCTCGGCGGCGCATGACATCTCGC 516
541 GTCACCTCGCTACGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
517 GTCACAGCTGACCGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
601 AACTTCTCAACGACGAGATCAACCCCATGCTCCCTTCGCGGCGGACATCTGCGGCTGCG 660

DB 577 AACTTCTCAACGACGAGATCAACCCCATGCTCCCTTCGCGGCGGACATCTGCGCTCG 636
QY 661 GAGGACCTTCTCCGCTCTCTCTAGATCGCGCGGCGCATCAACCGGTCAACCGGACGNNAG 720
DB 637 GAGGACCTTCT 696
QY 721 GTNCACTGNNNTNCAAGAGGCGANNAGAGATCATGNNCGCCGCGGAGTGGCTC 780
DB 697 GTGCACTGCTCAAGAGGCGAGAGAGATCTGTAGCGCCCGGAGGAGTGGCTC 756
QY 781 TTNAGTCTGAGCGCGGTGCTCTGCGCGCGGAGAGGAGTCTGAGTCTGCAACGCGACG 840
DB 757 TTCAACTCTGAGCGCGGTGCTCTGCGCGCGGAGAGAGTCTGAGTCTGCAACGCGAC 816
QY 841 GCGCTCTCGCTCGATGCGGACCCCTGCTCTGCAAGCAACATGCTCTGCTCTC 900
DB 817 GCGGCTCTGAGATGATGCGGACCCCTGCTCTGCAAGCAACATGCTCTGCTCTC 876
QY 901 TCGGAGGCGCTCAAGGCTCTTACAGTGAAGGCGATGTCGCGGCGGCGGCTGCTCAC 960
DB 877 TCGAGTCTCTGAGCGGCGGATCAAGTGAAGGCGATGTCGCGGCGGCGGCTGCTCAC 936
QY 961 CCNTTCTCAACGAGTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAGTGA 1020
DB 937 CCCTTCTTCAAGCGGCTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAG 996
QY 1021 CGGAGGCTCTGAGGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 997 GCGAAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
QY 1081 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
DB 1057 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
QY 1141 GCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB 1117 GCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1177 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1236
QY 1261 TTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
DB 1237 TTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296
QY 1321 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1297 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1356
QY 1381 CCNTTCTCAACGAGTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAGTGA 1440
DB 1357 CCCTTCTTCAAGCGGCTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAG 1416
QY 1441 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1417 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1476
QY 1501 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1477 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
QY 1561 AAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1537 AAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
QY 1621 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 1597 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1656
QY 1681 CCNTTCTCAACGAGTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAGTGA 1740
DB 1657 GCTGCGCTCAACGAGTCAAGCGGCGGCTTCAACCGGACCGAGTGAAGTGAAGTGAAG 1716

06-FEB-1987; 87GP-00024705.
PR 18-JUN-1987; 87GP-00152357.
XX (MITSU) MITSU TOATSU CHEM INC.
XX Fukuhiara N, Yoshino S, Sone S, Nakajima Y, Makiuchi N;
PI WPI, 1988-229543/33.
DR
PT Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT having combined promoter of tac promoter and pI promoter for enhanced
PT expression.
XX
XX Example; Fig 2(A)-2(D); 37pp; English.
XX
XX Plasmids were from E.coli MT 10414 (perm-8876) carrying the plasmid
CC pyrB contg. the structural gene for PAL from Rhodospiridium rubroide.
CC Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC for PAL (AA81101) is described as a reference example. The hybrid
CC plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC for the prodn. of L-phenylalanine from cinamic acid and ammonia (Updated
CC on 25-MAR-2003 to correct pr field.) (Updated on 27-MUG-2003 to correct
CC OS field.)
XX
XX Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;
SQ
Query Match 67.8% Score 1677.4; DB 1; Length 2331;
Best Local Similarity 80.4% Pred. No. 1.6e-225;
Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6

1 ATGGCCCCCTCCNTGCATCTGCATCCGCGACTCTCAGTCCGCCAACGGCCTCNCMAAGGTTNG 60
1 ATGGCACCTCGCTGCATCTGCATCTGCATCTGTTCCGAACGCGCTGCATCCGCACAAAG 60
61 CAGCGCCGNNCCGNCNNNGNCNCNACGCGGACCCAGTGCACNCTCNCNCGCGCGNCNG 120
61 CAGGCTGT-----CAATGGCCCTTCGACCAACTC-----GCAGTGGCAG 100
121 GCTCGCTCTCTCCGACACCCAGNNGACGACGACTCGACATGTNGAGNAGATCTCTGGCG 180
101 GCTCGACACTGGCCCAACACCCAGTGCACGACGAGTGCATGTGAGAAAGATGCTGGCG 160
181 ACCCGACCGNACGNAACGNNNTGAACTCGACGCGGTACACCTCACCTCGGNCACGTG 240
161 CGCGCACCGACTGCA---CGCTCGAACTCGACGCTACTCGCTTAACCTCGAGACGTG 217
241 TCGCGCGCGNCGCAGAGGGCGCGCNCNGTCCGCGTNCAGACAGNCGACGATCCGCGCA 300
218 TCTCGCGCGGAGAGGAGGCGACGCTGTCCGCGTCAAGAGCAG---CGACGAGATCCGCTCA 276
301 AAGATCGACAAAGNCTCGAGTCTCTCCGNNNCAGCTCUNACAACAGNCTACGAGTGC 360
277 AAGATTGACAAATCGTGTGAGTCTTGGCGCTGCACTCTTCAAGAGCGTCTACGCGTTC 336
361 ACGACTGATTTCGCGCGGCTCGGCGCACACCGGACTGAGATGCGNATCTCGCTTCAGAG 420
337 ACGACTGAGATTGGCGGATCCGACAGACACCGGACCGGAGCGGCATCTCGCTCAAGAG 386
421 GNCCTCTCGAGACCAAGCTTGGCGGTGTCCTCCMACGTCGAGNCTGAGTCTCTCCGCTC 480
397 GCTCTCTCGAGACCAAGCTTGGCGGTGTCCTCCCTTGTGTGACTGTTGCTGCTC 456
481 GAGCGCGGCTCGAGACTCGCTTCGCTCGAGTGTGCTCGGCGGCATGACCATTCGCG 540
457 GCGCGGGTCTGAGAACTCGTTCCTCCCTCGAGGTGTTGTCGCGCGCCATGACAAATCCGC 516
541 GTAACTCGCTCAACGCGCGGCACTTGGCGNCTCGGCTCTGTGTTCTTCAAGGCGCTTACC 600
517 GTCAACAGCTTGCACCGCGGCACTTGGCGTGTCCGCTCTGTGTTCTTCAAGGCGCTTACC 576
601 AACTCTCTCAACAGGAGATCAACCCCATGCGCCCTCCGCGGACCAATCTCGGCGCTCG 660

Db	577	AACTTCCTCAACGAGGATCACCCCATGTGCCCTCCGGGAGACATCTGCGCG	636
Qy	661	GGGGACTCTCCCNCTCTNTAATGCGCGGCCATACGGGTACCCGGACNNCAAG	720
Db	637	GGGACCTCTCTCTCTCTCTCAATGACGGGCATAGAGGTACCCGGACGCAAG	696
Qy	721	GNMCAAGTNMNCAGAGGGCANNAGAAATGATGNGGCCCGGAGGGCATGCGCGTC	780
Db	697	GTGCACTGTGTCCAGAGGGGCAGAGAAAGTCTGTACGCCCGGAGGGCATGTGCGCTC	756
Qy	781	TTNGGTCTGAGCCCGTGTCTTGCGGCCGAGAGAGGTCTGTCTGTCAACGGCAG	840
Db	757	TTCAACCTGAGCCCGTGTCTGTCTGCGCCGAGAGAGGTCTGGTGTGTCAACGGCAC	816
Qy	841	GCCGTCTCGCTCGATGGCAACCTTGCTGTGACAGACGACACATGTCTCGTCTC	900
Db	817	GCGGTTCAGATCATGTGGCCACCTCGCTGTGACAGACGACACATGTCTCGTCTC	876
Qy	901	TGCAAGGACCTCACGGCTCTNACGGTGAAGGCATGGTCGGACACCGCGCTGTCCAC	960
Db	877	TGCAAGTGTCTACGGGCATGACGGTGAAGCATGTGTGGCCACCGCGGCTGTCCAC	936
Qy	961	CNNTTCTTCAAGACTACGGGCTTACCCGACCCAGTGAAGTGTGCGCGCAATC	1020
Db	937	CCCTTCTTCAAGAGTCAACGGCGCCCTCACCCGACGATGCAATGCGGGGAAATC	996
Qy	1021	CGACAGCTCTGAGGGGAGGNNGTGTCGGTCCACACAGAGGAGAGGTCAAGTCAAG	1080
Db	997	CGCAAGCTCTGAGGGAGAGCGGCTTGCTGTCCACATGAGGAGAGGTCAAGTCAAG	1056
Qy	1081	GACGACGAGGGATTTCTCGCGACGACCGCTACCGCTCGACGTGCTCTAGTGTCTC	1140
Db	1057	GACGACGAGGGATTTCTCGCGACGACCGCTACCGCTTGACAGCTCTCTAGTGTCTC	1116
Qy	1141	GGCGCGCTGTGAGGACATGATTCAGCGCCACGNNGTCTCTCGTGTGAGGCGGNCAG	1200
Db	1117	GGCGCGCTGTGAGGACCTCATTCAGGCCACGGCGTCTTACCATGAGGCGGGCAG	1176
Qy	1201	TGACGACCGCAACCCGCTCATGACGTGTGAGAACAGANGAATCCACACGCGCGCAAC	1260
Db	1177	TGACGACCGCAACCCCTCTCATGACGTGTGAGAACAGACTTCGACACGCGCGCAAT	1236
Qy	1261	TTTCAGAGGCGCGGTGTGCGNAACAGATGAGAGATCTGCGTGTGCGNCTGCGCTGTGATC	1320
Db	1237	TTTCAGAGCTGCGCGTGTGACCAACAGATGAGAGATCTGCGTGTGCGGCTGTGCGCATATC	1296
Qy	1321	GGCAAGCTCAATTCACGACAGTCAACGAGATGTCAACGCGGCAATGAACGCGGCGCTN	1380
Db	1297	GGCAAGCTCAATTCACGACACTCAACGAGATGTCAACGCGGCAATGAACGCGGCGCTC	1356
Qy	1381	CNNTCTGTGTGTGTGCGGAGAACCNCTCGTGTCTCTATATCATGCAAGGGCTGCACATT	1440
Db	1357	CCCTCTGTGTGTGTGCGGAGAACCCCTGTCTCTCAACATGCAAGGCTGTCAACATC	1416
Qy	1441	GGCGNCGNCTACATTCATTCGAGAGTGGNACCTTCCAAACCGGTYNCAACCAACGTC	1500
Db	1417	GCGGTGTGCGGCTAACCTGTGAGATGTGAGACACTGTGCACACCTTGTGACGACGATGTG	1476
Qy	1501	CAGCGGAGAGATGGGCAACGAGCGGTCAACTCGTGTGNCNTAATCTGGACNCGCGC	1560
Db	1477	CAGCGGCTGAGATGGGCAACGAGCGGTCAACTCGTGTGNCNTAATCTGGCTGTGTGC	1536
Qy	1561	ACNCGGAGAGGCAAGAGTCTTCTCTCTCTGCGCACCACTCACTCACTGTGCTTC	1620
Db	1537	ACGACGAGATCAAGAGTCTTCTCTCTCTGCGCACCACTCTAATCTGTGCTTTC	1596
Qy	1621	CAGCGGTGACCTCGCGCGATGTGAGTTCAAGAGATTCAGACCCGANTNTC	1680
Db	1597	CAGGCAATGACCTTGCGCGGATGTGAGTTCAAGAGATTCGCGCCACGACATC	1656
Qy	1681	MCNNGCTCTNAGAGACGACTTTGGCNCNCCCTGACGCGNNNNNNNNNNNNAACAA	1740
Db	1657	GTCTGTCTATGACACGACTTTGTGTCTGCGCATGACGGCTGTGAACCTGTGCGACGAG	1716

QY 1742 CTCGAGGACCAAGGCTCAACAAAGCGCTCMAACAACGACTGAGGAGACCAACTCGTAGAC 1805
 Db 1717 CTCGTGCGAAGAGTGAACAAACACGCTCGCCAAAGCGCTTGAGCAACAACCTGTGACAC 1776
 QY 1801 CTCGAGCCGCGCTGGACACGACGCTTCTCGTTGGCACCGGACACGCTCGTCAGANNNNN 1860
 Db 1777 CTCGTCCGCGCTGGACAGACGCTTCTCTTTGGCCGCGGACCGCTGTCAGAGTCTCC 1836
 QY 1861 NNCTCTCGCCNNNNGCCANNAGTCTCGCTCGCAGCGCTCAAACGCTTGGAAAGTCCCT 1920
 Db 1837 TCGTCAGAGTC-----GCTTCGCTCGCCGCGCTCAACGCGCTGAGAGTCCCG 1885
 QY 1921 CGCGCGAAGAGGCACTTCGCTCAAGGCGMANGTCCGCAACGCTTCGAGCGCGCGT 1980
 Db 1886 CGCGCGAGTCGCGCACTCGCTCACCCGCAAGTCGCGAAGCTTCTGTCGCGCGT 1945
 QY 1981 CGTCGCTCGCGCCGCGCTCAGCTCTCGCTCGCGGCAAGCGCGCTCTGTACTGTCG 2049
 Db 1946 CAGACTGTGCGCGCGGCTCTGTACTCTGCGCGGCACTAGATCTCTAGCGCTTCC 2005
 QY 2041 TCCGCGAGGAGCTCGGCGTCAAGCGCCGCGCGGCGACGCTTCTTCGCGAAGGAGG 2100
 Db 2006 TCCGCGAGGAGCTTGGCTCAAGCGCCGCGCGGAGAGCTTCTTCGCGAAGGAGG 2065
 QY 2101 TCACGATCGGACCAAGAGCTCCCGCATCTACGAGGCAATCAAGNNCGCAGTCAAC 2166
 Db 2066 TCACGATCGGCTCGACACTTCCAAAGATCTACAGGCGCATCAAGTGGGCGAGATCAAC 2122
 QY 2161 AGCTCTCGTCAAGTGTCTGCGTATAGNNCNCNNNCNANNCTCGCNTNNNNCCNNNCNC 2220
 Db 2126 AGCTCTCTCTCAAGATGCTCGCTTAGACACTCTTCCCATCTCGCATCCCTTCATACC 2185
 QY 2221 NNCCNNNNNNNNCTTTNGNNNTCGNNCTGACNNNNCGAGANNNTNNCCNNNNNNN 2280
 Db 2186 TATCCGCGCTGCATCTTAGACTGCGCTCTGTGCGACTCGGATCTGCGATCGCTTCTT 2244
 QY 2281 TNNNNCTTCTCTNNCTCNCNNNNANCCNNTNNNNCTNNNNCTNNNNNNNNNNC 2340
 Db 2246 TCGTCTTTGGCTGCTCTTC-TAGACCGTGTGGATTAAGTGAATGGAATCAAGC 2300
 QY 2341 NGTNNNCANNA 2352
 Db 2305 AGTACCACATCA 2316
 RESULT 7
 ID AAN81116
 AC AAN81116 standard; DNA; 2331 BP.
 XX
 DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 12-NOV-1990 (first entry)
 DE L-phenylalanine ammonialyase structural gene.
 KM L-phenylalanine ammonialyase; PAL; foreign gene expression;
 KM culture temperature; expression regulation; ss.
 OS Rhodospiridium rubroideum; (IRO 559).
 FH Key Location/Qualifiers
 FT CDS 1. 2151
 FT /tag= a
 FT /Product= "PAL"
 PN EP279665-A.
 XX
 PD 24-AUG-1988.
 XX
 PF 18-FEB-1988 88EP-00301356.

XX 19-FEB-1987; 87JP-00034397.
 PR 18-JUN-1987; 87JP-00152359.
 XX (MITK) MITSUI TOATSU CHEM INC.
 PA
 XX Fukushima N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y;
 PI WPI; 1988-236895/34.
 DR P-PSDB; AAP80513.
 XX Regulation of expression of foreign gene in Escherichia coli - by
 PT maintaining temp. of culture at 40 deg. C or more to suppress expression.
 XX
 PS Disclosure;); 36pp; English.
 CC The cloned DNA from pSM13 isolated from the cloned cell carrying pSM13
 CC was digested with various restriction endonucleases. The sequences from
 CC the fragments were analysed by the method of Maxam-Gilbert and also by
 CC biochemical means using the method of Maat. The sequences obtained were
 CC edited using the GENAS program produced by Mitsui Information Development
 CC Co. The sequence so determined is represented here. The regulation of
 CC expression of this foreign gene in E.coli carrying a recombinant plasmid
 CC comprises maintaining the culture at at least 40 deg. C. See also EP-
 CC 279664. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
 CC 2003 to correct PI field.) (Updated on 24-OCT-2003 to standardise OS
 CC field)
 CC
 XX
 SQ Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;
 Query Match 67.8%; Score 1677.4; DB 1; Length 2331;
 Best Local Similarity 80.4%; Pred. No. 1.6e-225;
 Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;
 QY 1 ATGGGCCCCCTCCCTGCACTGATCGGACCTCGATCGGCAAGGCTGTCGAGGTTG 60
 DB 1 ATGGACCCCTCGCTGCACTGATCTTGCACTGTTTGCAAAAGGCGTGCACTCCGCAAG 60
 QY 61 CAGCGCCGNNCCGNNNGNCGNNAACGGGCGCCACGCTCAGCTCGNCGCGCCGCG 120
 DB 61 CAGGCTGT-----CATGCGCTCGACCAACCTC-----GCACTCGAG 100
 QY 121 GCTCGCTCTCCCGACCAACCGAGGAGGAGTGAATGCTGAGTGAAGTCTCTCCCG 180
 DB 101 GCTCGACCTGCGCCCAACCGAGGTCAGCAGAGTGAATGCTGAGTGAAGTCTCTCCCG 160
 QY 181 ACCCGACGNNACGNNACGNNNTGAACTGAGCGGCTACACCTCAGGAGCTG 240
 DB 161 CGCGACGACTGCA---CGCTGAACTGAGGCTACTCGCTCAACCTCGAGACGTCG 217
 QY 241 TCGGCGCCGCGCGCAAGGCGCGCGNCGTCCGCTGCAAGAGAGAGAGTCCGCGCA 300
 DB 218 TCTCGCGCGCGCAAGGCGCGCGTCTCGCTGCAAGAGAGAGAGTCCGCTGCA 276
 QY 301 AAGATGACAANAAGTGAAGTTCTCGGNNCGNCGCTCMAACAAGAGTCTACGAGTC 360
 DB 277 AAGATTGACAATAAGTGAAGTTCTTGAGCTGCAACTCTTCAAGAGGCTTACGAGCTC 336
 QY 361 ACGACTGATTTGGCGGCTCGGCGCAACCCGGAAGTGAAGTGCATCTCGCTCAGAG 420
 DB 337 ACGACTGATTTGGCGGATTCGCGACACCGCGCAACGAGAGCGCATCTCGCTCAGAG 396
 QY 421 GCGTCTCTGAGAGCAACAGTCTGCGGTGTCCTCCNAGTGAAGTCTCTGAGCTC 480
 DB 397 GCTCTCTGAGAGCAACAGTCTGCGGTGTCCTCTGCTGCTGAGTCTCTGAGCTC 456
 QY 481 GAGCGGCGCTCGAGAACTCGCTTCGCTCGAGGTCGTCGCGGCGCGCATGACATCGCG 540
 DB 457 GCGCGGCGCTCGAGAACTCGCTTCGCTCGAGGTCGTCGCGGCGCGCATGACATCGCG 516
 QY 541 GTCAACTGCTCAACGCGGCGCACTCGGCGTTCGCTGCTGCTTGAAGCGCTGACC 600
 DB 517 GTCAACAGCTGACCGCGGCGCACTCGGCTGCTGCTGCTGCTGAGGCGCTGACC 576

QY 601 AACTTCCTCAACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 DB 577 AACTTCCTCAACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 636
 QY 661 GCGGACCTCTCCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
 DB 637 GCGGACCT 696
 QY 721 GTNCACTGNNYACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
 DB 697 GTGCACTGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
 QY 781 TTNGGCTCGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 DB 757 TTCAACTCGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
 QY 841 GCGGCTCTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 DB 817 GCGGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 876
 QY 901 TCGGAGGCGCTCAAGGCTCTMAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
 DB 877 TCGAGTGTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
 QY 961 CCNTTCTCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 937 CCCTTCTCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
 QY 1021 CGGAGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
 DB 997 CGGAGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
 QY 1081 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 DB 1057 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
 QY 1141 GCGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 DB 1117 GCGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1176
 QY 1201 TCGAGAGCGCAACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1177 TCGAGAGCGCAACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1236
 QY 1261 TTCCAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
 DB 1237 TTCCAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1296
 QY 1321 GCGAGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 DB 1297 GCGAGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
 QY 1381 CCNTTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
 DB 1357 CCNTTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1416
 QY 1441 GCGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
 DB 1417 GCGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1476
 QY 1501 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 DB 1477 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
 QY 1561 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 DB 1537 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
 QY 1621 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
 DB 1597 CAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1656

QY 1681 NCNNGCTGCTNTNAGACGACTTTGGCNCNGCCTCGACGCGCANNNNNNNNNNACGAA 1740
DB 1657 GTCTCGCTCATGACAGACACTTTGGCTCCGCGATGACCGGCTCGAACCTCGCGAGAG 1716
QY 1741 CTCGNGGACGAAGGTACAAGCGCTCNAACAAGCGNCTCGAGGACCAACCACTGCTAGAC 1800
DB 1717 CTGTGGAAGAGTGAACAAGCGCTCGCAAGCGCTCGAGACCAACTGCTTAGAC 1776
QY 1801 CTGAGCCCGCGCTGCGACGACGCTTCTGTTCGGACCGGACCGCTGCTGAGANNNNN 1860
DB 1777 CTGTCCCGCGCTGGGACGCGCTTCTTCTTCGCGCGGACCGTGTGAGAGTCTC 1836
QY 1861 NNCTCTCCGANNNNGCCANNAGTCTGCTGTGCGNCGCTCAAGCGCTGGAAGTCCGCT 1920
DB 1837 TCGTCAAGCTC-----GCTCTGCTGCGCGCGCTCAAGCGCTGGAAGTCCGCG 1885
QY 1921 CGCGCAGAGAGGCACTCTGCTCAAGCGGANGTCCGCGACNCTTCTGAGCGGCGT 1980
DB 1886 CCGCGAGTGGGCACTCTGCTCAAGCGGANGTCCGCGACNCTTCTGAGCGGCGT 1945
QY 1981 CGTGTGTCGCGCGCGCTCNGCTACTCTGCGCGGACGCGCGCTCTGACTGTTG 2040
DB 1946 CGACCTGTGCGCGCGCTCTGTAAGCTCTGCGCGGACGCTGATCTCTACGCTTCG 2005
QY 2041 TCGCGAGAGAGCTCGGCGTCAAGGCGCGCGCGGAGGCTCTCTCGGCAAGGAGG 2100
DB 2006 TCGCGAGAGAGCTTGCGCTCAAGGCGCGCGCGGAGGAGCTCTCTCGGCAAGGAGG 2065
QY 2101 TGAAGTGGCAACAAGTCTCCGCACTCTAGAGCGCATCAAGNCGCGCATCAACC 2160
DB 2066 TGAAGTGGCTGCAAGTCTCTCAAGATCTAGAGCGCATCAAGTGGGCGAGATCAACA 2125
QY 2161 AGCTCTCTCAAGTGTCTGCTGANNNNNCGANNCTCGGCTNNNNNNNNNNNNC 2220
DB 2126 AGCTCTCTCAAGTGTCTGCTGAGCACTCTTCCACTCTCGGATCCCTTCATACCC 2185
QY 2221 NNN 2280
DB 2186 TATCCGCGCTGCACTTATGAGACTGCTTCTTGTGGAATCGGATCTCGCTTCTT 2245
QY 2281 TNN 2340
DB 2246 TCGTCTCTGCTGCGCTCTC-TAGACCGTGTGCTATTAAGATTGTGAATPACAGC 2304
QY 2341 NGTNNNNNNNN 2352
DB 2305 AGTACCATCA 2316

RESULT 8
ID AAN81399 strand; cDNA; 2328 BP.
AC AAN81399;
DT 12-DEC-1990 (fixer entry)
DE Sequence of Rhodospiridium toruloides IFO 559 cells cDNA encoding L-
DE Phenylalanine ammonia lyase (PAL).
XX Aspartame; artificial sweetener; enzyme; ss.
XX Rhodospiridium toruloides.
OS
FH Key Location/Qualifiers
FT CDS 1..2151
FT misc_feature 1..2148
FT /tag= b
FT /tag= b
FT /note= "Claimed"
XX
XX BP260919-A.

PD 23-MAR-1988.
XX 15-SEP-1987; 87EP-00308119.
PF 16-SEP-1986; 86EP-00215864.
PR (MITR) MITSUI TOATSU CHEM INC.
XX Fukuhara N, Yoshino S, Yamamoto K, Se T, Sone S, Nakajima Y;
PI Suzuki M, Makiguchi N;
XX MPI: 1988-079223/12.
DR P-PSDB; AAB81099.
PT New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for
PT producing L-phenylalanine from cinamic acid and ammonia.
PS Disclosure; 9; 37bp; English.
CC PAL is useful in prodn. of L-phenylalanine from cinamic acid and an NH3
CC donor (process claimed). L-Phe is an essential AA and starting material
CC for sweetener aspartame. E.coli transformants will produce PAL without
CC induction by expensive AAs; no surfactant is needed to increase cell wall
CC permeability
XX
XX Sequence 2328 BP; 423 A; 889 C; 591 G; 425 T; 0 U; 0 Other;
SQ
Query Match 67.7%; Score 1675.8; DB 1; Length 2328;
Best Local Similarity 80.4%; Pred. No. 2.7e-225;
Matches 1890; Conservative 0; Mismatches 426; Indels 36; Gaps 6;
QY 1 ATGGGCCCCCTCCNTGACTGATGCGGACCTCGMTGCGCAAGCGCTCNCMAACGAGTNG 60
DB 1 ATGGGACCCCTGCTGACTGATGATCTGCACTGTTCGAAAGGAGTGTGCTCCGAAAG 60
QY 61 CAGGCGNNNCCGNN 120
DB 61 CAGGCTGT-----CAATGGCGCTCGACCAACTC-----GCACTCCGAG 100
QY 121 GCTGCTCTCTCCCGACCAACCAAGNAGAGCTGAGATGTGAGNAGATCTCTGCGCG 180
DB 101 GCTGAGCTTCCCGCAACCAAGGTCAGAGGTGATGTGAGAGATGTCTGCGCG 160
QY 181 ACCCGACGNNNAGAACNNNTGAACTCGACGAGGTACACCTCACTCGGAGCGTGC 240
DB 161 CGCGACGACTCGA---CGCTGAACTCGAAGGCTACTGCTCACTCGGAGACGTCG 217
QY 241 TCGGCGCGCGNCGCAAGGCGCGCNCNCTCGCGCTGTCAGACAGNCGAGATCCGCGCA 300
DB 218 TCTCGGCGCGCGAGAGGCGAGGCTGTCCGCTGTCAGAGAG-CGAGGATCCGCTCA 276
QY 301 AAGATCGACAAAGAGTGAAGTCTCTCCGANNCGAGCTCNAACAAGTCTACGAGTGC 360
DB 277 AAGATTGACAAATCGGTGAGTCTTGTGCGCTCGAACTCTCATGAGGCTTACGGGCTC 336
QY 361 ACGACTGTTTCGGGCGCTCGGCGCACACCGGACTGAGATGAGTGTGCTCCAGAG 420
DB 337 ACGACTGATTTCGGGAGTTCGACAGACACCGGACCGGACGCGCATCTCGCTCAGAG 396
QY 421 GGNCTCTGAGACACACTCTGCGGCTGTCNCCNAGTGGANTGANTCTTCCGCTC 480
DB 397 GCTCTCTGAGACACACTCTGCGGCTTCTCTCTGCTGTGTGATCTGTTCGCTC 456
QY 481 GAGCGGCGCTCGAAGACTCGTCTCGCTCGAGTGTGCGCGGCGCATGACCATCGCG 540
DB 457 GCGCGGCTCTGAGAACTCGTCTCTCGAGGTGTGCGGCGCATGACATCGCG 516
QY 541 GTCAAGTCTGACNCGGCGGCACTCGGCGTCTGCTGTCTGTCTGAGGCGCTCAAC 600
DB 517 GTCAAGCTGATCCCGGCGGCACTCGGCTGTGCGCTGTGCTGTGAGGCGCTCAAC 576
QY 601 AACTTCTCAACGAGGATACCCCACTGATCCCGCTCGCGGACCATCTCGGCGTGC 660

QY 1681 MCNNGCTGTCNAGAGGACCTTGGGCGCCCTGAGCGGNNNNNNNNNNNACGAA 1740
Db 1708 CCGACTCTCTCCAGAGGACCTCGGCGCTGGCTGAGCTC-----AACCA 1755
QY 1741 CTGNGGACAGAGGTCACAAAGCGCTCNAAGCGNCTGAGAGACCACTGTAAGAC 1800
Db 1756 CTGCGCTCGAGGTCAAGAGGCGCTCAACAGCGCTGAGACAGACGTAAGAC 1815
QY 1801 CTGAGCGCGCTGAGACAGCGCTTCTGTTGCGAGCGGACCGCTGAGNNNNN 1860
Db 1816 CTGAGCGCGCTGAGACAGCGCTTCTGTTGCGAGCGGACCGCTGAGCTCTC 1875
QY 1861 NNNGCTCTCGCGNNNNGGANNAGTCTGCTGCGGCGGCTGAAGCGTGAAGGCT 1920
Db 1876 TCGTCTCTCGCTCTGCGCA--AGTCACTCTTAATGCGCTCAACGCGTGAAGGCT 1933
QY 1921 CCGCGGAGAGGCGCTCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
Db 1934 CCGCGGAGAGGCGCTCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1993
QY 1981 CGTCTGCTGCGCGCGCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
Db 1994 CTGCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2053
QY 2041 TCGCGGAGAGGCTCGGCGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
Db 2054 TGGCGGAGAGGCTCGGCGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2113
QY 2101 TGAAGTGGGACCAAGGCTTCCCGATCTAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
Db 2114 AGAGGATCGGAGCAAGGCTTCCCGATCTAGAGGCGGCGGCGGCGGCGGCGGCGG 2173
QY 2161 ACCTCTCTGAGAGGCTCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
Db 2174 ACCTCTCTGAGAGGCTCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2232
QY 2221 NNNCGNNNNNNCTTNGNNNTGNNNTGNNNTGNNNTGNNNTGNNNTGNNNTGNNNT 2280
Db 2233 ACCCGAAGACGAGCTTTCGAGCTCGTCTGAGCAAGAGCGGCTTCTCAATACACA 2292
QY 2281 TNNNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNN 2340
Db 2293 TGTGCTTACTCTCTGCGGCTGATACGCTCTCAAGTTCTTCTGATCCGCGCTCTCTC 2352
QY 2341 NGTNNNCAANNACNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400
Db 2353 GGTGCTCA-GTACAGCTGTATAGAGCTGGAATGATTCGAAGTCTTCAAGTCAAAAA 2411
QY 2401 NNNNAAAA 2408
Db 2412 AAAAAA 2419

RESULT 10
AAD45812
ID AAD45812 standard; DNA; 2419 BP.
XX
AC AAD45812;
DT 27-DEC-2002 (first entry)
XX
DE Rhodotorula graminis PAL DNA #1.
XX
KM Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
KM human cytomegalovirus infection; cytostatic; virocid; gene; de.
XX
OS Rhodotorula graminis.
XX
FH Key Location/Qualifiers
FT CDS 37..2199
FT /*tag= a

FT /product= "PAL protein"
FT /EC_number= "4.3.1.5"
FT /transl_except= (pos:493..495, aa:Xaa)
FT /*note= "Xaa corresponds to Val or Ala"
US002102712-A1.
01-AUG-2002.
24-AUG-2001; 2001US-00939408.
24-JUL-2000; 2000US-00624693.
24-JUL-2001; 2001WO-US023270.
PCBU SERVICES INC.
Yoshida RK, Koetsira AB;
WPI; 2002-690616/74.
P-PSDB; AAE27938.
Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
polypeptide, useful for treating a mammal having phenylketonuria, cancer,
human immunodeficiency virus or human cytomegalovirus infection.
Claim 13; Page 27-29; 74pp; English.
The present invention relates to Yeast (e.g. Rhodotorula) phenylalanine
ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
such proteins. PAL sequences are useful for producing L-phenylalanine.
They are useful for treating mammals having diseases, disorders or
conditions that would benefit from treatment with PAL proteins such as
phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
human cytomegalovirus (hcmv) infection. Sequences of the invention are
also used in gene therapy. The present sequence is R. graminis PAL DNA
Sequence 2419 BP; 416 A; 909 G; 659 G; 434 T; 0 U; 1 Other;
Query Match 65.8%; Score 1629; DB 6; Length 2419;
Best Local Similarity 78.0%; Pred. No. 9.1e-219;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
QY 1 ATGGCCCTCTGCTGAGTCTGATCGGACCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db 37 ATGGCCCTCTCTGAGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 96
QY 61 CAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 97 CAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 154
QY 121 GCTGCTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 155 GCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 214
QY 181 ACCCGAAGACGAGCTTTCGAGCTCGTCTGAGCAAGAGCGGCTTCTCAATACACA 240
Db 215 ACCCGAAGACGAGCTTTCGAGCTCGTCTGAGCAAGAGCGGCTTCTCAATACACA 271
QY 241 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 272 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 330
QY 301 AAGATCGACAAAGAGTCTGAGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 331 CGGCTGACAAAGAGTCTGAGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
QY 361 ACGAGTGGTTTCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 391 ACGAGGTTTCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450
QY 421 GCGTCTCTGAGACCAAGCTTGGGCTGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 451 GCGTCTCTGAGACCAAGCTTGGGCTGTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 510

QY	481	GGNCCGCGCTCGAGAACTCGCTTCGCGCTGAGAGTGGTCCGCGCGCCATGACATCGCG	540
Db	511	GGACCGGCGCTCGAGAAACCGCTTCGCGCTGAGAGTGGTCCGCGCGCCATGGTATCGCG	570
QY	541	GTCCAACTGCTCAACCCGCGCTCACTCGGCGGTCCGCGCTGGTGGTCTCTCGAGGCGCTCAAC	600
Db	571	GTCAACTCGCTCAACCGGTGGCCACTCGGCGGTCCGCGCTGGTGGTCTCTTGAAGCGCTCAAC	630
QY	601	AACTTCCTCAACCAAGGATCAACCCCATGGTCCCGCTCGCGGACCAATCTCGGCGTCCG	660
Db	631	AACTTCCTGAACACCGGATCAACCCCATGGTCCCGCTCGCGGCTCCATCTCGGCGTCCG	690
QY	661	GGCGACTCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG	720
Db	691	GGCGACTCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCG	750
QY	721	GTNCAAGTNTNCAAGAGGAGCANNAGAAATATATGNTGCGCGCGGAGCGATCGCGCTC	780
Db	751	GTTCACGTTTTCAGAGAGGAGCAGAAATATATGTTTCGCGCGAGGCGCATCTCGCTC	810
QY	781	TTNGGCTTCAGAGCCGCTGGTCTTCGCGCGGAGAGGAGTTCGAGTCTCGTCAACGCGCACG	840
Db	811	TTTGGTCTCGAGGAGATGGTCTCTCGCGCGGAGAGGAGTTCGAGTCTCGTCAACGCGCACG	870
QY	841	GCGGTCTCGCTCGATGGCGACCTTCGCTTCGACGACGACACATGCTTCGCTCTC	900
Db	871	GCGGTCTCGCTCGATGGCGACCTTCGCTTCGACGACGACACATGCTTCGCTCTC	930
QY	901	TCGCAAGGCGCTCAAGGCTCTNACGATCGAGGCGCATGGTTCGCGCACGCGCGGCTGTTTCAC	960
Db	931	TCGCAAGGCGCTCAAGGCTCTNACGATCGAGGCGCATGGTTCGCGCACGCGCGGCTGTTTCAC	990
QY	961	CCNTTCCTTCAGAGCTCAAGGCTTCACCCGACCCGATCGAGGTTCGCGCGCAATC	1020
Db	991	CCGTTTCATCAAGAGTTCGCGCGCGCGCACCCGCGAGGTTCGAGGTTCGCGCGCAATC	1050
QY	1021	CGGAGGCTTCGAGAGGAGCGGNTTTGCGCTCCACAGAGGAGAGGTTCAGGTTCAAG	1080
Db	1051	CGGAGGCTTCGAGAGGAGCGGNTTTGCGCTCCACAGAGGAGAGGTTCAGGTTCAAG	1110
QY	1081	GACGAGAGGAGGATTCCTCGCGCAGAGCGGCTACCGCTTCGCGACGTTCGCTCAAGTGGCTC	1140
Db	1111	GACGAGAGGAGGATTCCTCGCGCAGAGCGGCTACCGCTTCGCGACGTTCGCTCAAGTGGCTC	1170
QY	1141	GCGCGCTTCGTCAGGACATGATTCAGGCCACGCGNETCCTTCGCTCGAGGCGCGGACG	1200
Db	1171	GCGCGCTTCGTCAGGACATGATTCAGGCCACGCGNETCCTTCGCTCGAGGCGCGGACG	1230
QY	1201	TCGAGGACCGGACCAACCGGCTCATGAGCGTCGAGGAGCAAGNAGCCACGACGCGCGCAC	1260
Db	1228	ACGACGACCGGACCAACCGGCTTCGAGCGTTCGAGGAGCAAGNAGCCACGACGCGCGCAC	1287
QY	1261	TTTCAGGCGNCGCTTCGCGCNAACAGATGAGAGAGCTTCGCTTCGCGCTTCGCGCTTCGCTTC	1320
Db	1288	TTTCAGGCGNCGCTTCGCGCTTCGCGCTTCGAGTTCGAGTTCGCGCTTCGCGCTTCGCTTC	1347
QY	1321	GGGAACTCACTTCACGAGCTCAACCGAGATGCTCAAGCGCGGACATGAAACCGCGGCTTN	1380
Db	1348	GGGAACTCACTTCACGAGCTCAACCGAGATGCTCAAGCGCGGACATGAAACCGCGGCTTN	1407
QY	1381	CCNTTCCTTCGCTTCGCGAGGAGCCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC	1440
Db	1408	CCNTTCCTTCGCTTCGCGAGGAGCCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC	1467
QY	1441	GCCGCGNCGNCTCACTTCGAGGCTTCGAGNCACTTCGCAACCGGTTNAGCACCACTC	1500
Db	1468	CACATGCTTCGCTTCGAGGCTTCGAGGCTTCGAGGCTTCGAGGCTTCGAGGCTTCGAGGCTTC	1527
QY	1501	CACGCGGAGATGGGACACGAGCGCTCAACTGCTCGCNCCTCATCTTCGCGNCGCGC	1560
Db	1528	CACGCGGAGATGGGACACGAGCGCTCAACTGCTCGCNCCTCATCTTCGCGNCGCGC	1587

QY		1561	CNCGCCGAGGCACAACAGCCTTCCTTCTCTCTCTCTGGCACCAACTCATCGGCTMCTC	1628				
Dd		1588	ACTGCCAGGSCAACAGACTCTTTCTCTCTCTCTGTCCGACACTGTATCGACGCTC	1647				
QY		1621	CAGGCGCTGCATCTCCGCGCGATGAGATTGATCAGAAGCAAGTTGACCCTNTNTC	1680				
Dd		1648	CAGGCGCTGCATCTCCGCGAGATGAGACTTGATCAGAAACAGTTGACCTCCCTCTC	1707				
QY		1681	NNNNNGCTCMTCMGACGACCTTTGGCNMGCCCTGAGCGGNNNNNNNNNNAACAA	1740				
Dd		1708	CCGACTCTCTCCACGACGACTCTGGCATTGGCCTTGACGTC-----AAGCCA	1755				
QY		1741	CTCGNGACAAAGTCAA CAAGNCGCTTCMA CAAAGCNTCTGAGCAAGCOACTCGTAGAC	1800				
Dd		1756	CTTGCGCTCGAGGTCAAGAAAGCGCTCAACAGCGCTTGACGAACAGACAGACGTACAC	1815				
QY		1801	CTCGAGCCGCGCTGGCACAGAGCCCTTCTGTTCCGACCGGACCCGTGTGAGNNNNNM	1860				
Dd		1816	CTCGAGCCGCGCTGACAGAGCCCTTCTGTAAGGACACCGGACCGGTGTGAGCTCTC	1875				
QY		1861	NNGTCTCTGC CNNNNGCCANNAGGCTTCGCTCGCNCCTCTCAAGCTTGAAAGTGGCT	1920				
Dd		1876	TGTCCTCTGCGCTCTGGCA--AGGTACCTTTATCTGCTGCAAGCGGTGAAAGTTGCTT	1933				
QY		1921	CCGCCGAAAGGCAATCTCGCTCAAGGCGANATCCGCACTCTTGTGANGGNCCTT	1980				
Dd		1934	CGCGGAAAGGCATCTCGCTCAAGGCGAGGTGGCAACCGCTTCTGGCAGAGCGCT	1993				
QY		1981	CGTCTCTGCGCCGCGCTCMGCACTCTCGCCGCGACGCGCTCTGTACTGTTGG	2040				
Dd		1994	CTTCCAGGCCCGGCGCACGCAATCTCTCGCGGCGACGGCGCTGTACTGTTGG	2053				
QY		2041	TCCGAGAGACTCGCGCTCAAGGCCCGCGCGACGCTTCTCTCGCAAGAGAGG	2100				
Dd		2054	TGCGGAGAGACTCGGCGTGGAGGCGGCGCGCACGCTGTGTCGCGCGAGAGG	2113				
QY		2101	TGACGATGGGAGCAACGCTCTCCCGCATCTACAGGCGCATCAAGNCGGCGCATCAAC	2160				
Dd		2114	AAGCATGGGAGCAAGCTCTCGCGCATCTACAGGCGCATCAAGAGCGCGCATCAAC	2173				
QY		2161	ACGCTCTGTCAAGATGCTGCGNTAGNNNCNCCNNNCNANCCTCGCNTNNNNNCNCC	2220				
Dd		2174	AGCTCTGTCAAGATGCTGCGATA-AGGCCGAGCAAGCTCGCTAGACCGCGCTC	2233				
QY		2221	NNNNNNNNNNCTNTTNGANTCGANTCTPNCNCCNNNCGGANNTTNNCNNNNNNNN	2280				
Dd		2233	ACCCAGACAGAGCTTTTCAGCGCTGCTGTGTCACAAAGCAGCTTCTCTCATACACA	2299				
QY		2281	TNNNNCTNNCTNNCTCNCCNNNNANCNNGTCCNTNNNNCTNNNGTNNNNNNNNNC	2340				
Dd		2293	TGTGCGCTACTCTCTGCGCGTATACAGCTCTCTGAGTCTTTTGTGATCCGCGTCTGC	2355				
QY		2341	NCTNNNCANNNA CNCTNNNNNNNANNCNANNGANNGANTTNAGNNTTNGNNGNNNN	2400				
Dd		2353	GGTCTCTCA-GTACACGTGTATAGAGCTGGAAAGATGTCAAGTCTTCAAGTTCAAAAA	2411				
QY		2401	NNNANAAA 2408					
Dd		2412	AAAAAAAA 2419					
RESULT 11								
AAD33076								
ID	AAD33076 standard; DNA; 2151 BP.							
AAD33076;								
01-JUL-2002 (first entry)								
Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.								
KW	Para-hydroxycinnamic acid; PHA; food packaging; electronic connector;							
KN	phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;							

KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
LM LCP; telecommunication; medical device; aerospace application; enzyme;
KW biocatalyst; gene; mutant; variant; ds.
OS Rhodotorula glutinis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 1..2151
FT CDS /product= "Rhodotorula glutinis mutant PAL"
W0200210407-A1.
XX
PD 07-FEB-2002.
XX
PF 23-JAN-2001; 2001WO-US002099.
XX
PR 27-JUL-2000; 2000US-00627216.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Gateby AA, Sarielani FS, Tang X, Qi WW, Vannelli T;
XX
XX WPI; 2002-280635/32.
DR P-PSDB; AAE20654.
XX
PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT polypeptide, or mutant TAL polypeptide, useful for producing
PT parahydroxycinnamic acid in recombinant host cell lacking cinamamate
PT hydroxylase.
XX
XX Example 10; Page 95-98; 139pp; English.
XX
CC The invention relates to methods for biological production of para-
CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC biocatalyst for conversion of glucose to PHCA by incorporation of the
CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC Rhodotorula glutinis into Escherichia coli, underlying the ability of the
CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC to developing a new biocatalyst for conversion of glucose to PHCA by
CC incorporation of the wildtype PAL which possesses enhanced tyrosine
CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC recombinant techniques. The recombinantly produced PHCA may be used as a
CC monomer for production of liquid crystal polymers (LCP). LCP may be used
CC in electronic connectors and telecommunication and aerospace
CC applications. LCP resistance to sterilizing radiation has also enabled
CC these materials to be used in medical devices as well as chemical, and
CC food packaging applications. The present sequence is Rhodotorula glutinis
CC PAL mutant DNA
XX
XX Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
Best Local Similarity 83.7%; Pred. No. 2.1e-218;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGGACCCCTCGTCACTGATCGAGCTGATCGCAAGCGGTCNCAACGAGTNG 60
DB 1 ATGGACCCCTCGTCACTGATCGAGCTGATCGCAAGCGGTCNCAACGAGTNG 60
QY 61 CACGCCGNNCCGNNNGCNCACGGGCGCCACGTCACNCTGNGCGCGCGCNG 120
DB 61 CACGCCGNNCCGNNNGCNCACGGGCGCCACGTCACNCTGNGCGCGCGCNG 120
QY 121 GCTGCTCTCCGACACCCAGNNNGACGAGCTGACATGAGTNGAGTCTCTCCG 180
DB 101 GCTGCTCTCTCCGACACCCAGNNNGACGAGCTGACATGAGTNGAGTCTCTCCG 180
QY 181 ACCCCACCGNNAAGNACGNNNTGAACTTGACGGGTACACCTCACTCGGAGAGCTG 240
DB 181 ACCCCACCGNNAAGNACGNNNTGAACTTGACGGGTACACCTCACTCGGAGAGCTG 240

DB 161 CGCCGACGCACTCGA---CGCTCGAAGTCGAGGCTACTCGCTCACTCGGAGAGCTCG 217
QY 241 TCGGCGCGCGCNCGCAAGGGCCGCGCNCGCTGTCNCAGACAGNCGACGAGATCGCGCA 300
DB 218 TCTCGGCGCGCGCAAGGGCGAGGCGCTGTCGCGCTCAAGAGACAG-CGACGAGATCGCTCA 276
QY 301 AAGATCGACAANAAGTGTGAGTCTCGGNNCNCAGCTCNAACAAGNGCTACGAGTGC 360
DB 277 AAGATTCGACAAATTCGATGAGTCTTGCGCTCGCACTTCATGAGCGCTTACGGCGTC 336
QY 361 ACGACTGATTTGCGCGCTCGCGCGACACCCGCACTGAGATGAGATGAGTCTCGTCAAG 420
DB 337 ACGACTGATTTGCGCGCTCGCGCGACACCCGCACTGAGATGAGATGAGTCTCGTCAAG 396
QY 421 GGNCTCTCGAGACACCACTCTGGGCTGTNTCTCCNACGTCGNTGANTCTTNGCTTC 480
DB 397 GCTCTCTCGAGACACCACTCTGGGCTGTNTCTCCNACGTCGNTGANTCTTNGCTTC 456
QY 481 GGNCGGCGCTCGAGAACTCGCTTCGCTCGAGGTCGTCGCGGCGCATGACATCGCG 540
DB 457 GCGCGCGCTCGAGAACTCGCTTCGCTCGAGGTCGTCGCGGCGCATGACATCGCG 516
QY 541 GTCAACTGCTGACACGCGCGCACTCGGCGTCTCGCTCTGTCGAGGCGCTCAC 600
DB 517 GTCAACAGCTTGACCCGCGGCACTCGGCTGTCCGCTCTGTCGAGGCGCTCAC 576
QY 601 AACTTCTCAACGACGAGATCACCCCATCGTCCCGCGGCACTCTGCGGCTG 660
DB 577 AACTTCTCAACGAGATCACCCCATCGTCCCGCGGCACTCTGCGGCTG 636
QY 661 GCGACCTCTCCGCTCTCTGATGAGCGCGCGCATGACCGGTCACCCGCAACG 720
DB 637 GCGACCTCTCTCTCTCTCTCTATGACGCGCGCATGACCGGTCACCCGCAACG 696
QY 721 GTCACGNNNNACGAGGAGGANNAGATATGNNCCCGCGGAGGATCGGCTC 780
DB 697 GTGACGCTGTCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
QY 781 TTGAGTCTGAGCGCGCTGCTCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
DB 757 TTCAACTCTGAGCGCGCTGCTCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 816
QY 841 GCGCTCTCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
DB 817 GCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 876
QY 901 TCGAGGCGCTCAAGGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
DB 877 TCGAGGCGCTCAAGGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 936
QY 961 CCGTCTCTCAAGGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
DB 937 CCGTCTCTCAAGGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 996
QY 1021 CGGAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 997 CGGAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
QY 1081 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
DB 1057 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
QY 1141 GCGCGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
DB 1117 GCGCGCTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
QY 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1177 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1236
QY 1261 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
DB 1237 TTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296

Qy	1321	GGCAAGCTCAATTACGACACTCACGAGATGCTCAAGCGGGATGAAACGCGGCTT	1380
Db	1297	GGAAAGCTCAATTACGACACTCACGAGATGCTCAAGCGGGATGAAACGCGGCTT	1356
Qy	1381	CCNTCTGCTCGCTGCGGAGAACCCNTGCTCTCTATCACTGCAGAGGCTCTGACATT	1440
Db	1357	CCCTCTGCTCGCGGGCGAAGACCCCTCGCTCTCTTACACTGCAGAGGCTCTGACATT	1416
Qy	1441	GGCGGAGGCTCACTTGGAGCTGGAGCACTTGGCAACCCGGTNAAGACCAAGTC	1500
Db	1417	GGCGCTGGGGGTACACTTGGAGCTGGAGCACTTGGCAACCCGGTNAAGACCAAGTC	1476
Qy	1501	CAGCGCGGAGATGGGCAACAGGCGGTCAACTGCTGCGNCTCATCTCGAGCCGCGC	1560
Db	1477	CAGCGCGGCTGAGATGGGCAACAGGCGGTCAACTGCTGCGNCTCATCTCGAGCTGCGC	1536
Qy	1561	ACNCGCGAGGCCAAGAAGTCTTCTTCTCTCTCTGCGCAACCACTCTACTGCGTNTC	1620
Db	1537	ACACCGAGTCCAAAGAGTCTTCTCTCTCTGCGCAACCACTCTACTGCGTNTC	1596
Qy	1621	CAGCGCGTGAAGCTCGCGCATGAGTGGAGTTCAGAGAGAGTTCAGCCGATNTTC	1680
Db	1597	CAAGCATGACTTTCGCGGACCGAGTTCAGAGTTCAGAGAGAGTTCGCGCCACATTC	1656
Qy	1681	NCNCGCTCTNACAGACACTTTGGGCGNCGCTCGACGCGNNNNNNNNNNNAAGAA	1740
Db	1657	GTCTGCTCATGACACAGACCTTTGGCTTCGCGCATGACGCGCTCGAACCTGCGCACAG	1716
Qy	1741	CTCGNAGCAAGAGTCAACAAGNCGCTCNACAGCGNCTCGAGAGACCACTGTACAGC	1800
Db	1717	CTCGTCGAAAGTGAACAGACCTTCGCAAGGCTTCGAGAGACCACTGTACAGC	1776
Qy	1801	CTGAGCGCGCTGGCAAGACGCTTCTGTTTGCGACCGGACCGTGTGAGNNNNN	1860
Db	1777	CTGCTGCCCGCTGGCAAGACGCTTCTGTTTGCGACCGGACCGTGTGAGGTCCTC	1836
Qy	1861	NNGTCTCGCCNNNNGCCANNAGGCTCTGCTCGCGCGCTCAACGCTGTGAGTGTGCT	1920
Db	1837	TGCTGAGCTC-----GCTCTGCTCGCGCGCTCAACGCTGTGAGTGTGCTG	1885
Qy	1921	CGCGCGAGAGGCTCATCTGCTCACGCGGAGNAGTTCGCGACNCTTTCTGAGCGGNCCT	1980
Db	1886	CGCGCGAGAGGCTCATCTGCTCACGCGGAGNAGTTCGCGACNCTTTCTGCGCGGCT	1945
Qy	1981	CGTGTGCTGCGCGCGGCTCNCGTACCTCTGCGCGGAGCGGCTGCTACTGTTG	2040
Db	1946	CGAGCTGTGCGCGCGGCTCTGTAAGCTCTGCGCGGAGCTCAAGATCTCTAGCGTTG	2005
Qy	2041	TCCGCGAGAGCTCGGCTGAGGCGCGCGCGCGCGAGCTTCTTCTGCGGACGAGG	2100
Db	2006	TCCGCGAGAGCTTGGCTGAGGCGCGCGCGCGCGAGAGCTTCTTCTGCGGACGAGG	2065
Qy	2101	TGACGATCGGAGCAAGAGTCTCCGAGCTTACGAGGCAATCAAGNNCGGCGCATCAAC	2160
Db	2066	TGACGATCGGAGCAAGAGTCTTCAAGATCTTACGAGGCAATCAAGTGGGAGAGATCAAC	2125
Qy	2161	ACGTCTCTGTCAGATGCTGCGGNTAG	2186
Db	2126	ACGTCTCTCTCAAGATGCTGCGTAG	2151
RESULT 12			
AAD33100 standard; DNA; 2151 BP.			
AAD33100			
AAD33100 standard; DNA; 2151 BP.			
AAD33100			
AAD33100			
01-JUL-2002 (first entry)			
R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, Epi8Km-6.			
Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;			
TW			

KW	phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KM	chromatophore p-450; cytochrome p-450 reductase; liquid crystal polymer;
KV	LCP; telecommunication; medical device; aerospace application; enzyme;
KW	biochemical; mutant; variant; gene; ds.
XX	
OS	Rhodotorula glutinis.
OS	Synthetic.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2151
FT	/tag= a
FT	/product= "Rhodotorula glutinis PAL mutant"
FT	/EC_number= "4.3.1.5"
FT	replace(645, G)
FT	/tag= b
FT	replace(792, A)
FT	/tag= c
FT	replace(858, T)
FT	/tag= d
FT	replace(1619, T)
FT	/tag= e
XX	
PN	MO200210407-AI.
XX	
PD	07-FEB-2002.
XX	
PF	23-JAN-2001; 2001WO-US002099.
XX	
PR	27-JUL-2000; 2000US-00627216.
XX	
PA	(DUFO) DU PONT DE NEMOURS & CO E I.
XX	
PI	Gatenby AA, Sarasiani FS, Tang X, Qi WW, Vannelli T,
DR	WPI; 2002-280635/32.
DR	P-PADB; AAE20668.
PT	Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT	polypeptide, or mutant TAL polypeptide, useful for producing
PT	parahydroxycinnamic acid in recombinant host cell lacking cinmmate
PS	hydroxylyase.
XX	
PS	Disclosure; Page; 139pp; English.
CC	The invention relates to methods for biological production of para-
CC	hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC	biochemical process for conversion of glucose to PHCA by incorporation of the
CC	wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC	Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC	wildtype PAL to convert tyrosine to PHA. The invention is also directed
CC	to developing a new biochemical process for conversion of glucose to PHCA by
CC	incorporation of the wildtype PAL which possesses enhanced tyrosine
CC	ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC	plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC	Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC	recombinant techniques. The recombinantly produced PHCA may be used as a
CC	monomer for production of liquid crystal polymers (LCP). LCP may be used
CC	in electronic connectors and telecommunications and aerospace
CC	applications. LCP resistance to sterilizing radiation has also enabled
CC	these materials to be used in medical devices as well as chemical, and
CC	food packaging applications. The present sequence is Rhodotorula glutinis
CC	PAL mutant DNA. Note: This sequence is not shown in the specification,
CC	however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC	(AAD30705) in the sequence listing
XX	
SQ	Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
XX	
Query Match	65.7%; Score 1626.4; DB 6; Length 2151;
Best Local Similarity	83.7%; Pred. No. 2.le-218;
Matches 1829; Conservative	0; Mismatches 322; Indels 35; Gaps 5;
XY	1 AATGGCCCCCTTCAGTCGATCGCATGCCACTCGTGTCCAAAGGCNTCNCMAACGGNNNG 60

Db 1 ATGGACCCCTCGCTCGACTCGATCTTGCACTGCTTCGCAAGCGGCTGCAATCCGCAAG 60
QY 61 CACGCCGNNCCGNCNNNGNCNMACGGGCGCCACGTCACNCTGNCNGCCGNCNG 120
Db 61 CAGGCTGT-----CAATGCGCCTCGACCAACCTC-----GAGTCGAG 100
QY 121 GCTCGCTCTCCCGACCAACCAAGNNGACGACACTGACATCGTTNAGNAGTCTCTCGCG 180
Db 101 GCTCGACCTGCCCAACCCAGTTCAGCAGGTGACATGCTGAGAGATGCTCGCG 160
QY 181 ACCCAACGNNACGNNNTGAACTCGAGGGTACACCTCACCCTCGANAGCTCG 240
Db 161 CGCCGACCGACTGA---CGCTGAACTCGAGGCTACTCGTCAACCTCGAGACGTCG 217
QY 241 TCGGCGCCGNCGCAAGGGCCGNCNGTCCGCTGNCAGACAGNCGACGAGATCCGCGCA 300
Db 218 TCTCGGCGCGGAGGAGGAGGCGCTGTCCGCTGAGAGAG- -GACGAGATCCGCTCA 276
QY 301 AAGATCGACAAAGNGTCCAGTTCCTCCGNNCAGCTCNACACAGNGTCTAGCGNGTC 360
Db 277 AAGATTGACAAATCGGTCAGATTCTTGCGCTCGCAACTCTCATGAGGCTTACGGGTC 336
QY 361 AGCACTGTTTCGGGCGCTCGGCGCACACCGGACTGAGATGCMATCTGCTCCAGAAG 420
Db 337 ACGACTGGAATTTGGCGGATCCGACAGACCCGCAACGAGAGCCATCTCGCTCAAGAAG 396
QY 421 GNCCTCTCGAGACAGAGCTTCGCGGTGTCNNCCGNNAGTCGNTGANTCTTNGCTC 480
Db 397 GCTCTCTCGAGACAGAGCTTCGCGGTGTCNNCCGNNAGTCGNTGANTCTTNGCTC 456
QY 481 GNCGCGGCGCTCGAGAACTCGCTTCGCTCGAGTGTGTCGCGCGGCGCCATGCCATCCG 540
Db 457 GCGCGGCGTTCGAGAACTCGCTTCGCTTCGAGTGTGTCGCGCGGCGCCATGCCATCCG 516
QY 541 GTCACTCGCTCAGNCGGCGCCACTCGGCGNCTCGCTTCGCTTCGAGGCGCTCAC 600
Db 517 GTCAACAGCTTGACCGCGGCGCACTCGGCTGTCGCTTCGCTTCGAGGCGCTCAC 576
QY 601 AACTTCCTCAACACGAGGCTCAACCCCATGTCCTCCGCGGACCATCTCGGCGTCG 660
Db 577 AACTTCCTCAACACGAGGCTCAACCCCATGTCCTCCGCGGACCATCTCGGCGTCG 636
QY 661 GCGGACCTCTCCCNCTCTCTNTACATGCGCGCGCCATCAACGCTTACCGGACNNCAAG 720
Db 637 GCGGACCT 696
QY 721 GTNCACTNNNTNCAAGAGGCGANNAGAAAGATCATGTCGCGCGGCGGAGCGATCGGCTC 780
Db 697 GTGACAGTGTTCACAGAGGCAAGAGAAAGATCTGTACGCGCGGAGGCGATGCGCTC 756
QY 781 TTNGTCTCGAGCCGCTGCTCTCGGCGCGGAGAGGAGTCTCGTTCACAGGACG 840
Db 757 TTCAACTCGAGCCGCTGCTCTCGGCGCGGAGAGGAGTCTCGTTCACAGGACG 816
QY 841 GCGGCTCTCGGCTCGATGCGGACCTCGCTCTGACAGAGCAACATGCTTCGCTCTC 900
Db 817 GCGGCTCTCGAGCATGATGCGGACCTCGCTCTGACAGAGCAACATGCTTCGCTCTC 876
QY 901 TCGAGAGGCTCAAGGCTCTNAGGCTCGAGGCGCATGTCGCGGCAAGCGGCTGTTCCAC 960
Db 877 TCGAGTCTCTCAAGGCGCATGAGGCTCGAGGCGCATGTCGCGGCAAGCGGCTGTTCCAC 936
QY 961 CCNTTCTCTCAAGAGCTCAAGCGGCTTACCCGACCGAGATGAGGTCGCGCGCAATC 1020
Db 937 CCCTTCTCTCAAGAGCTCAAGCGGCTTACCCGACCGAGATGAGGTCGCGCGCAATC 996
QY 1021 CGAGGCTCTCGAGGCGAGGNNGTTCGCGGCAACCGAGAGGAGGCTCAAGGTCGAG 1080
Db 997 CGGAGCTCTCGAGGAGGCGGCTTGTGTGTCACATGAGAGAGGAGTCAAGGTCGAG 1056
QY 1081 GACGAGAGGAGGATCTTCGCGGAGAGCGCTACCCGCTCGGACGTCGCTCACTGAGTGC 1140
Db 1057 GACGAGAGGAGGATCTTCGCGGAGAGCGCTACCCCTTGGCGCATGTCCTCACTGAGTGC 1116

QY 1141 GGCCCGCTGTGTACGACATGATTCAGGCCACGNCNTCTCTCTGCTCGAGGCGCGNCAG 1200
Db 1117 GGCCCGCTGTGTACGACATGATTCAGGCCACGNCNTCTCTCTGCTCGAGGCGCGNCAG 1176
QY 1201 TCGAGACCGACCAACCGCTCATGAGCTGAGAGCAAGANAGACCGACGAGCGGCGCAC 1260
Db 1177 TCGAGACCGACCAACCGCTCATGAGCTGAGAGCAAGACTTGTGACACAGGCGGCAAT 1236
QY 1261 TTCAAGCGGCGCTGTGTGNCNAAACGATGAGAGAGACTCGGCTCGNCCTCGCTGATC 1320
Db 1237 TTCAAGCGGCGCTGTGTGNCNAAACGATGAGAGAGACTCGGCTCGGCGCTCGGCGATC 1296
QY 1321 GCGAAGCTCAACTTCAAGGAGCTCAAGGATGTCGACGCGGAGTGAACCGCGGCTG 1380
Db 1297 GCGAAGCTCAACTTCAAGGAGCTCAAGGATGTCGACGCGGAGTGAACCGCGGCTG 1356
QY 1381 CCNTCTGCTCGCTGCGGAGGACCCNTGCTCTCTATCATGCAAGGCGCTCGACAT 1440
Db 1357 CCCTCTGCTCGCTGCGGAGGAGACCCCTCGCTCTCTATCATGCAAGGCGCTCGACAT 1416
QY 1441 GCGGCGGCGGCTTACATTCGAGCTGAGNACCTTGCCAAACCGGTCAGACCGTC 1500
Db 1417 GCGGCTGCGGCGTACCTCGAGTGTGAGACCTCGCCAAACCTGTGAGAGCGCATGTC 1476
QY 1501 CAGCGGCGGAGATGAGGCAACGAGCGGTCACCTCGTGCNCTCATCTCGGCGGCGG 1560
Db 1477 CAGCGGCTGAGATGAGGCAACGAGCGGTCACCTCGTGCNCTCATCTCGGCGGCGG 1536
QY 1561 ACNCGGAGGCGCAACGAGTCTTCTCTCTCTCTGCGCACCGACCTCTACGCTGCTC 1620
Db 1537 ACGACGAGTCCAAAGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGGCGGCTGAGCTCGGCGGATGAGTGAAGTGAAGAGAGTTCGACCGGNNNC 1680
Db 1597 CAGGCTATGAGCTTGGCGGCGACCGAGTGTGAGTTCAGAGAGAGTTCGCGCGACATC 1656
QY 1681 NCCNCGCTCNCNAGCAGCACTTTGGCNCGCCCTCGACCGGNNNNNNNNNNAGAA 1740
Db 1657 GTCTGCTCATGACAGCACTTTGGCTCCGCGCATGACCGGCTGCAACCTGCGCGACGAG 1716
QY 1741 CTGNGAGCAAGGTCACCAAGNCGCTNACAGGNCNTCGAGAGACCACTGTGAGC 1800
Db 1717 CTGTCGAGAAAGGTGAACAGACGCTGCGCAAGGCGCTCGAGAGACCACTGTGAGC 1776
QY 1801 CTGAGCGGCGCTGSCAGACGACCTTCTCGTTGCGGACCGGACCGGCTGCGAGNNNNN 1860
Db 1777 CTGTCGCGGCTGSCAGACGACCTTCTCTCTGCGCGGCGACCGGCTGCGAGTCTC 1836
QY 1861 NNGTCTTGCCNNNNGCCANNAGTCTGCTCGGCGGTCGCAACGCTGGAAGTGCCT 1920
Db 1837 TCGTGAAGTC-----GCTTCGCTCGCGCGGTCGCAACGCTGGAAGTGCCTG 1885
QY 1921 CCGCGAGAAAGGCTATGCTGTCAGCGGCAAGTCCGCGGACNCTTCTGNCGNCCT 1980
Db 1886 CCGCGAATGCGGCTATGCTGTCAGCGGCAAGTCCGCGGACNCTTCTGTCGCGCGCT 1945
QY 1981 CGTGCCTGCGCGGCGCTGNCGATCTGCGCGGCGGACGCGGCTGCTTACTGCTTCG 2040
Db 1946 CGACTGCTGCGCGGCGCTGNCGATCTGCGCGGCGGACGCGGCTGCTTACTGCTTCG 2005
QY 2041 TCCGAGAGAGCTGAGGTCAGAGCGCGCGGCGGAGCTTCTCTCGGCAAGAGAG 2100
Db 2006 TCCGAGAGAGCTGAGGTCAGAGCGCGCGGCGGAGCTTCTCTCGGCAAGAGAG 2065
QY 2101 TGACGATGAGCAACGATCTCCGCGCATCTACAGAGGCGCATCAAGNNCGGNCGATCAAC 2160
Db 2066 TGACGATGAGCTCAAGCTCTCAAGATCTACAGAGGCGCATCAAGTGGGCGAGATCAAC 2125
QY 2161 AGCTCTCTCAAGATGCTCGGCTAG 2186
Db 2126 AGCTCTCTCAAGATGCTCGCTTAG 2151

RESULT 13
 AAD26924
 ID AAD26924 strand: DNA; 2151 BP.
 AC AAD26924;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL) DNA.
 XX
 KW Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
 KW para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
 KW liquid crystal polymer; LCP; electronic connector; telecommunication;
 KW aerospace application; chemical; red yeast; mutant; ds.
 XX
 OS Rhodotorula glutinis.
 XX
 OS Synthetic.
 FH Key Location/Qualifiers
 FT CDS 1..2151
 FT /tag= a
 FT /product= "R. glutinis mutant PAL enzyme"
 FT mutation replace(1619, T)
 FT /tag= b
 XX
 PN WO200111071-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 03-AUG-2000; 2000WO-US021156.
 XX
 PR 06-AUG-1999; 99US-0147719P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Tang X, Vannelli TM, Qi WW, Sariaslani S, Garebby AA;
 XX
 PI WPI: 2002-121549/16.
 DR P-PSDB; AAEI6390.
 XX
 PT Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
 PT comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia-
 PT lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
 PT activity.
 XX
 PS Claim 13; Page 67-68; 755P; English.
 XX
 CC The invention relates to a method for producing para-hydroxycinnamic acid
 CC (PHCA). The method comprises converting cinnamate to PHCA by converting
 CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
 CC route or by generating a new biocatalyst possessing enhanced tyrosine
 CC ammonia-lyase (TAL) activity. The method is useful for the biological
 CC production of PHCA which is useful as a monomer for the production of
 CC liquid crystal polymers (LCP), where the LCP is used in electronic
 CC connectors, telecommunications and aerospace applications. LCP is also
 CC useful in medical devices, as well as chemical, and food packing
 CC applications due to its resistance to sterilizing radiation. The present
 CC sequence is a DNA encoding red yeast, Rhodotorula glutinis
 CC (Rhodospirillum toruloides) mutant PAL derived by substituting Ile at
 CC position 540 with Thr. Mutant PAL enzyme possesses enhanced TAL activity
 XX
 SQ Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
 Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 2.1e-218;
 Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
 QY 1 ATGGCCCTTCCTCGACTCGATCGGACCTCGTGGCAAACGCGTCTGCAACGAGTGG 60
 DB 1 ATGGACCTCTCGCTCGACTCGATCTCGCACTCGTTCGGCAAAGGGGTGCGATCGCAAG 60
 QY 61 CAGCCGNNCCGACNNNGCNCNACGAGGGCGGACGCTCCACNCTCNGNGCGCGNGCG 120

DB 61 CAGGCTG-----CAATGGGCGCTTCGACCAACCTC-----GCAGTCGAG 100
 QY 121 GCTGCTCTCTCCCGACCAACCCAGNNAGCGAGCTCGACATCTGTTGAGNAGATCTCTGCCG 180
 DB 101 GCTGCACTCTCCCAACAACCAAGGTACAGAGGTGACATCTGTGAGAAATCTCTGCCG 160
 QY 181 ACCCAACCGNNAACNAGCANNNTGAACTCGACGAGGTACACCTTCACCTCTGAGAGCTCG 240
 DB 161 CGCCGACCGACTCGA---CGCTGAACTCGACGCTACTGCTCACTGCACTTGAGAGCTCG 217
 QY 241 TCGGCGCGCGNCGAGAGGGCGCGNCTGCCGCTCNCAGACAGNCGACGAGATCCCGCA 300
 DB 218 TCTGCGCCCGCAGAGAGAGGACGAGCTCTCCGCTCAAGAGCAG--CGACGAGATCCGCTCA 276
 QY 301 AAGATCGACAANNAAGTTCGAGTTCTCTCGNNAACAGCTCAACACAGNCTCTACGNGTC 360
 DB 277 AAGATTGACAATTCGCTGAGTTCTTCTGCGCTCGCAACTCTTCATTCAGCGCTACGCGCTC 326
 QY 361 ACGACTGGTTTGGGGGCTGGGCGGACACCGGACCTGAGATGCAATCTGTGCTCCAGAAG 420
 DB 337 ACGACTGGATTGGCGGATCGGACAGACCCGACCGAGAGAGCCGATCTGTGCTCCAGAAG 396
 QY 421 GGNCTCTCGAGCAACGACTCTGCGGTGTCCTCCNACAGTCTGATGCAATCTTCCGCTTC 480
 DB 397 GGTCTCTCGAGCAACGACTCTGCGGTGTCCTCCNACAGTCTGATGCAATCTTCCGCTTC 456
 QY 481 GGNCGGCGCTCGAGAACTCGCTTCCGCTGAGAGTGTCCGCGCGCGCATGACATCTCCG 540
 DB 457 GCGCGGCTCTCGAGAACTCGCTTCCGCTGAGAGTGTCCGCGCGCGCATGACATCTCCG 516
 QY 541 GTCAACTGCTCACNCGGCGGCACTCGGCGNGTCGCGCTGCTCGAGGCGCTCAC 600
 DB 517 GTCAAGACTGTAACCCGCGGCACTCGGCTGTGCGCTCTGCTCGAGGCGCTCAC 576
 QY 601 AACTCTCTCAACCAACGCGATACCCCATGTGCTCCCTCCGCGGCAACATCTGCGCTCG 660
 DB 577 AACTCTCTCAACCAACGCGATACCCCATGTGCTCCCTCCGCGGCAACATCTGCGCTCG 636
 QY 661 GCGGACCTCTCCGCTCTGTCATGTCGCGCGGCACTCAACCGGTCACCGGCAACNCAAG 720
 DB 637 GCGGACCT 696
 QY 721 GTNCACTGTTNNTNCAAGAGGCGANNAGAAATCATGTNCGCCCGGAGGCGATCGCGCTC 780
 DB 697 GTCAGCTGTCGACAGAGGCAAGAGAAATCTTACGCGCGGAGGCGATGCGCTC 756
 QY 781 TTNGCTTCGAGCGCGTGTCTCTCGGCCCAAGAGGCTCTCGTCTCTCTCTCTCTCTCTCT 840
 DB 757 TTCAACTTCGAGCGCGTGTCT 816
 QY 841 GCGCTCTCGGCTCGATGCGACCTCGCTCTGACAGCAGACATGCTCTCTCTCTCTCTCTCT 900
 DB 817 GCGCTCTGAGATGATGCGACCTCGCTCTGACAGCAGACATGCTCTCTCTCTCTCTCTCTCT 876
 QY 901 TCGAGGCGCTTCAGGCTCTTACGCTGAGGCAATGCTGCGCCACGCGGCTCTCTCTCTCTCT 960
 DB 877 TCGAGCTGCTCACCGCATATGCTGCAACGATGCTGCGCCACGCGGCTCTCTCTCTCTCTCT 936
 QY 961 GNTTCTCTCAAGATGACGAGCGGCGCTCACCGGACCGAGATGAGGTGCGCGCAATC 1020
 DB 937 CCTTCTCTCAAGATGACGAGCGGCGCTCACCGGAGATGAGGTGCGCGCAATC 996
 QY 1021 CGACGCTCTCTGAGGAGCGAGNNGTTTGCCTTCACCAACGAGAGAGGTCAAGTCAAG 1080
 DB 997 GCGAAGCTCTCTGAGAGGAGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1056
 QY 1081 GACGACGAGGATCTCTCGGACAGACCGCTACCGGCTCTCGGACAGTCTCTCTCTCTCTCTCT 1140
 DB 1057 GACGACGAGGATCTCTCGGACAGACCGCTACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1116
 QY 1141 GCGCGCTCTCTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200

Db 1117 GGCCTGCTGTCAGCAGCTTCATTCAGCCGACGCGCTTCACCATCGAGGCGGCGAG 1176
 QY 1201 TCGACGACCGAGCAACCCGCTCATCGACGTCGAGGAACAGANAGACCACACGCGGCGAAC 1280
 Db 1177 TCGACGACCGAGCAACCCGCTCATCGACGTCGAGGAACAGATTCGACACGCGGCGCAT 1236
 QY 1261 TTCAGGCGGCGCTGCTGCGAACAAGATGAGAAAGACTCGCTGCGCTGCGCTGATC 1320
 Db 1237 TTCAGGCTGCGCTGCTGCGAACAAGATGAGAAAGACTCGCTGCGCTGCGCTGATC 1296
 QY 1321 GGCAGACTCACTTCAGCAGCTCACCGAGATGCTCAACGCGGCTATACCGGCGCTT 1380
 Db 1297 GGCAGACTCACTTCAGCAGCTCACCGAGATGCTCAACGCGGCTATACCGGCGCTC 1356
 QY 1381 CCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 Db 1357 CCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1416
 QY 1441 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 1417 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1476
 QY 1501 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Db 1477 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1536
 QY 1561 ACNCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 1537 ACNCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1596
 QY 1621 CAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 Db 1597 CAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1656
 QY 1681 NCNCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Db 1657 GTCGCTCATCGACACGACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1716
 QY 1741 CTCGAGCAAGGTCACCAAGGCTCNAAGGCTGAGCAAGGCTGAGCAAGGCTGAGCAAG 1800
 Db 1717 CTCGAGCAAGGTCACCAAGGCTCNAAGGCTGAGCAAGGCTGAGCAAGGCTGAGCAAG 1776
 QY 1801 CTCGAGCGGCTGAGCAAGGCTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 Db 1777 CTCGCTCCGCGCTGAGCAAGGCTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1836
 QY 1861 NNCTCTGCGCANNNGCCANNAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 Db 1837 TCGTCGACGCTC-----GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1885
 QY 1921 CCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 Db 1886 CCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1945
 QY 1981 CCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 Db 1946 CCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2005
 QY 2041 TCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 Db 2006 TCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2065
 QY 2101 TCGAGCTGCGGACCAAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
 Db 2066 TCGAGCTGCGGACCAAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2125
 QY 2161 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2186
 Db 2126 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151

RESULT 14
 ABS56543

ID ABS56543 standard; DNA; 2151 BP.
 AC ABS56543;
 XX 27-JAN-2003 (first entry)
 DT
 XX
 DE R. glutinis DNA encoding PAL/TAL naturally occurring variant.
 XX
 KM ds; gene; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase;
 KM PAH; para-hydroxycinnamic acid; tyrosine ammonia lyase; TAL; PAL/TAL;
 KM tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
 XX
 OS Rhodotorula glutinis.
 XX
 FH Key Location/Qualifiers
 FT 1. 2151
 FT CDS /tag=a
 FT /product= "PAL/TAL"
 FT
 PT WO200290523-A2.
 XX
 PD 14-NOV-2002.
 XX
 PF 03-MAY-2002; 2002WO-US018551.
 XX
 PR 04-MAY-2001; 2001US-0288701P.
 XX
 PA (DUPO) DU POINT DE NEMOURS & CO E. I.
 XX
 PI Qi WW, Sarielani FS, Tang X;
 XX WPI; 2003-058928/05.
 DR P-PsDB; AB671255.
 XX
 PT Novel recombinant host for producing para-hydroxycinnamic acid and
 PT tyrosine, comprises gene encoding tyrosine ammonia lyase activity and
 PT gene encoding phenylalanine hydroxylase activity.
 XX
 XX Example 4; Page 66-67; 69pp; English.
 PS
 XX The invention relates to a recombinant host comprising at least one gene
 CC encoding a tyrosine ammonia lyase (TAL) activity, and at least one gene
 CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
 CC providing the recombinant host; and (2) growing the recombinant organism
 CC in the presence of a fermentable carbon substrate and thus producing para
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
 CC providing a recombinant organism comprising at least one gene encoding a
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant
 CC organism in the presence of a fermentable carbon substrate to produce
 CC tyrosine. The recombinant host cell is used for producing para-
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
 CC monomer for the production of liquid crystal polymer. The present
 CC sequence encodes a naturally occurring variant of the R. glutinis
 CC Phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL).
 XX
 SQ Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
 Query Match 65.7%; Score 1626.4; DB 7; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 2.1e-218;
 Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
 QY 1 ATGGCCCCCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 Db 1 ATGGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 QY 61 CAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 Db 61 CAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100
 QY 121 GCTGCTCTCTCCGACCAACCAAGGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

Db 101 GCTGGACCTGCGCCACAACCCAGGTCAAGCAAGTGCATGTCGAGAAGTGTGCGCG 160
Qy 181 ACCCCACCGGNNNAAGAACGNNNTGCAACTTCGACGGGTACACCTTACCTCGGAGACGTG 240
Db 161 CCGCCAGCCGACTCGA---CGCTCGAACTCGAAGGTACTCTGCTCACTCGGAGACGTG 217
Qy 241 TCGGGCGCGCNGCGAAAGGCGCGCNCNCTGCGGTGTCAGACAGNCGAGAGATCCGCGCA 300
Db 218 TCTGGGCGCGGAGAAAGGCGAGGCTGTGTCGGGTGTAAGAGACAG--CGACGAGATCCGTCA 276
Qy 301 AAGATCGACAAANAGNGTGAATTCTCCGANNCAAGTCAAAAGNGTCAAGNGTC 360
Db 277 AAGATTGACAAATCGGTGAGTTCTTGCGCTCGCAACTCTCATGAGGTCGTAAGCGGTG 336
Qy 361 AGCATGAGTTTGGGGGCTCGGCGCAACCCGGAGTGAAGATGCAATTCGCTCGAAG 420
Db 337 AGCATGAGTTTGGGGGATCCGAGACACCCGGACCGAGAGCGCATTCCTCGTCAAG 396
Qy 421 GGNCTCTCGAGACCAAGCTTGCAGTGTCTCCGNNCAAGTCAGTCAATCTTCCGCTC 480
Db 397 GCTCTCTCGAGACCAAGCTTGGGGTGTCTCCCTTGTGCTTGAAGTGTGCGCTC 456
Qy 481 GGNCGCGCTCGAGAACTGCTTCCGCTCGAGGTGTGTCGGGGCGCATGACATCCG 540
Db 457 GCGCGGCTCTGAGAACTGCTTCCCTCGAGGTGTGTCGGGGCGCATGACATCCG 516
Qy 541 GTCAATCGCTCAACNGCGGGCGCACTCGGNGTCGCGCTGTGTCGAGAGCGCTCAC 600
Db 517 GTCAAGACTTGAACCGCGGCGCACTGAGTGTGTCGGCTGTGTCGAGAGCGCTCAC 576
Qy 601 AACTTCTCAACCAAGCGCATCAACCCCATGCTTCCCTCGCGGACCATTCGCGGTG 660
Db 577 AACTTCTCAACCAAGCGCATCAACCCCATGCTTCCCTCGCGGACCATTCGCGGTG 636
Qy 661 GGGGACTTCTCCGNNCTCTGTAACATCGCGCGCGCATCAACGTCACCCGGAAG 720
Db 637 GGGGACTTCTCTCTCTCTCTCAATGCAAGCGCATCAAGGGTCAACCCGAGCAAG 696
Qy 721 GTNCAGTNNNCAAGAGGAGAAATCATGTGCGCGCGAGAGCGATCGCGTC 780
Db 697 GTGACAGTGTCAAGAGGAGAAAGATCTGTACGCGCGGAGCGATGAGCGCTC 756
Qy 781 TTNGATCTGAGGCGCTGCTCTCGCGCGGAGAGGCTTCTGTCGTAAGCGACG 840
Db 757 TTCAACTCGAGGCGGCTGCTCTCGCGCGGAGAGGCTGCTGTCGTAAGCGACG 816
Qy 841 GCGGTCTCGGCTCGATGAGCGACCTCTGCTGTCAGACAGCAATGCTCTGCTCTC 900
Db 817 GCGGTCTGAGCATGATGAGCAACCTCTGCTGTCAGACAGCAATGCTCTGCTCTC 876
Qy 901 TCGCAGGCGCTCAAGGCTCTNAAGGTGAGGCGCATGTGCGCAAGCGCGCTGTTCCAC 960
Db 877 TCGCAGTCTGTCAGGCGCATGAGGTGAGAGATGATGTGCGCAAGCGCGCTGTTCCAC 936
Qy 961 CCGTCTCTTCAGAGTCAAGCGGCTTCAACCCGACCCAGATGAGGTGCGCGCAATC 1020
Db 937 CCGTCTCTTCAGAGTCAAGCGGCTTCAACCCGACCCAGATGAGGTGCGCGCAATC 996
Qy 1021 CGAAGCTCTCGAGGCGAGGNNGTGCGTCAACAGAGAGAGAGTCAAGGTCAAG 1080
Db 997 CGAAGCTCTCGAGGAGAGCGCTTGTGTCACATGAGAGAGAGTCAAGGTCAAG 1056
Qy 1081 GACGAGAGGAGATTTCTCGCGAGAGACGCTACCGGCTCGAGGTGCTTCAAGTGTG 1140
Db 1057 GACGAGAGGAGATTTCTCGCGAGAGACGCTACCGGCTTGTGCAAGTGTGCTG 1116
Qy 1141 GCGCGGCTCTGAGCGACATGATTCAGAGCGCAAGGNGTCTCTGCTCGAGGCGGAG 1200
Db 1117 GCGCGGCTCTGAGCGACCTTCAAGCGCGCAAGCGGCTTCAAGTGTGAGGCGGCGAG 1176
Qy 1201 TCGAGCAGCAGACCGGCTCATGAGGTGAGAGCAAGAGACCAAGCGCGCGCAAC 1260
Db 1177 TCGAGCAGCAGACCGCTCTCATGAGGTGAGAGCAAGACTTTCGAGCAAGCGCGCAAT 1236

Qy 1261 TTCCAGGCGCGGCTGTGCAACAGCATGAGAGAGACTGCGCTCGGCTCGCTGATC 1320
Db 1237 TTCCAGGCTGCGGCTGTGAGCCAAACCATGAGAGAGATCGGCTCGGAGCTGCGCAATC 1296
Qy 1321 GCGAAGCTCAACTTCACGAGCTCAAGAGATGCTCAACCGCGCATGAACCGCGCTT 1380
Db 1297 GCGAAGCTCAACTTCACGAGCTCAAGAGATGCTCAACCGCGCATGAACCGCGCTC 1356
Qy 1381 CCGTCTGCTCTGCTGCGGAGAGACCGTCTGCTCAATGAGTCAAGAGGCGTCAATT 1440
Db 1357 CCGTCTGCTCTGCGGCGGAGAGCCCTGCTCTCTCAACATGAGAGGCGTCAATC 1416
Qy 1441 GCGGAGGAGGAGTCACTTCGAGACTCGAGACTTTCGCAACCGCGTNAAGACCAAGTC 1500
Db 1417 GCGGCTGCGGAGTACCTGAGAGTGGGACCTTCGCAACCTGTGAGAGAGCATGTC 1476
Qy 1501 CAGCGCGAGAGATGGGCAACAGGCGCTCACTGCTGCGTCAATCTGCGGCGCGC 1560
Db 1477 CAGCGCGTGAATGGCAACAGGCGGTCACTGCTGCGTCAATCTGCGTCAATCTGCG 1536
Qy 1561 ACGGCGAGGAGCAAGAGCTCTTCTCTCTCTCTCGGCAACCGCTCTACTGCGTCTC 1620
Db 1537 AGGACGAGTCCAGAGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
Qy 1621 CAGGCGTCACTTCGCGGAGATGAGTTCAGATTCAAGAGAGTTCAGACCGGNTTTC 1680
Db 1597 CAAGCATGACTTCGCGGAGCAAGTTCAGATTCAAGAGAGTTCAGGCGGAGCATC 1656
Qy 1681 NCGNCGCTCTNAGACCACTTTGGGCGGCGCTTCGACCGGCGGNNNNNNNAGAA 1740
Db 1657 GTCTGCTATGACCAACACTTGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716
Qy 1741 CTCGAGCAAGAGTCAACAGGCGCTCNAAGGCGCTCAGAGAGCACTGTAAGAC 1800
Db 1717 CTCGAGAGAGTGAACAGAGCTTCGCAAGGCGCTCAGAGAGCACTGTAAGAC 1776
Qy 1801 CTCGAGCGGCTGCGAGAGGCTTCTGTTGCGGACCGGCAACCGTGTGTAAGNNNN 1860
Db 1777 CTCGTCGCGCTGCGAGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
Qy 1861 NNGTCTGCGGCGGNNNCCANNAGTCTGCTCGGCGGCGGCGGCGGCGGCGGCG 1920
Db 1837 TCGTCAAGCTC-----GCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCG 1885
Qy 1921 CCGCGAGAGGCGCATCTGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1980
Db 1886 CCGCGAGTGGGCGCATCTGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1945
Qy 1981 CGTGTGCTGCGCGGCGGCTCAGCTCTCTGCGGCGGCGGCGGCGGCGGCGGCG 2040
Db 1946 CGACTCTGTCGCGGCGGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2005
Qy 2041 TCGGAGAGAGCTCGGCGTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2100
Db 2006 TCGGAGAGAGCTTGGCGTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2065
Qy 2101 TGAAGTGGGCAACAGTCTTCGCGCATTCAGAGCATCAAGNNCGGCGGCGGCG 2160
Db 2066 TGAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2125
Qy 2161 AGTCTCTCTCAAGATGCTCGGCTAG 2186
Db 2126 AGTCTCTCTCAAGATGCTCGGCTAG 2151

RESULT 15
ADE36055
ID ADE36055 standard; DNA; 2151 BP.
XX ADE36055;
XX
DT 29-JAN-2004 (first entry)

QY 1321 GGCAAGCTCAACTTCAACGAGCTCAACGAGATGCTCAACGCGGCGATGAACCGGGCTTN 1380
DB 1297 GGCAAGCTCAACTTCAACGAGCTCAACGAGATGCTCAACGCGGCGATGAACCGGGCTTC 1356
QY 1381 CCNCTCTGCTCGCTGCCGAGAGCCNCTGCTCTCTATCACTGCAAGGGCTTGACATT 1440
DB 1357 CCTCTCTGCTCGCGGCGGAGAGCCCTCGCTCTCTACCACTGCAAGGGCTTGACATC 1416
QY 1441 GCGCGNCGNCTTAACCTTGGAGTCCGAGNACCTTGGCAACCGGTTNAGACCCAGCTC 1500
DB 1417 GCGGCTGCGGCTTACCTCGAGTTGGAGACCTTGGCAACCTTGGAGACGAGATGTC 1476
QY 1501 CAGCCGCGNAGATGGGCAACGAGGCGCTCAACTGCTCGNCTCATCTCGGCGCGC 1560
DB 1477 CAGCGGCGTGAATGCGGCAACGAGGCGTCAACTGCTTGGCTCATCTCGGCTGTCG 1536
QY 1561 ACNCGCGAGGCCCAAGAGTCTTCTCTCTCTCTCGGCAACCACTTACTGTGNTCTC 1620
DB 1537 ACGACGAGATCCACGAGAGTCTTCTCTCTCTCTCTCGGCAACCACTTACTGTGNTCTC 1596
QY 1621 CAGGCGGTGACCTCGGCGCGATGAGTTGAGTTCAAGAGCAGTTGACCCGNTNNTC 1680
DB 1597 CAAGCCATGACTTGGCGGCGGACCGAGTTGAGTTCAAGAGCAGTTGAGCCGACCATC 1656
QY 1681 NCNCGCTCTNTGACGAGCACTTTGCGNCGNCGCTCGACGCGNNNNNNNNNNNACGAA 1740
DB 1657 GTCTGCTCATGACGAGCACTTTGGCTCCGCAATGACGGCTCGAACTTGCGGACGAG 1716
QY 1741 CTCGNGAGCAAGGTCAACAAGNCGTCTNAACAAGNCTCGAGCAGACCACTCGTAGAC 1800
DB 1717 CTCGTGAGAAAGGTAAACAAGCGTCTGCAAGCGCTCGAGCAGACCACTCGTAGAC 1776
QY 1801 CTCGAGCCGCGCTGGGACGAGCCCTTCTGCTTGGCGGACCGGACCGCTGTCGAGNNNNN 1860
DB 1777 CTGTGTCGCGGCTGGGACGAGCCCTTCTGCTTGGCGGACCGGACCGGTGTGAGGTCTC 1836
QY 1861 NNGTCTCGCCNNNNGCCANNAGTCTGCTGCGNCGGCTCAAGGCTGGAAGTCTGCT 1920
DB 1837 TCGTCGACCTC-----GCTCTGCTGCGCGCGCTCAAGCGCTGGAAGTCTGCG 1885
QY 1921 CCGCGGAGAAAGGCACTCTGCTCAAGCGGNNANGTCGCGGACNCTTGGNCGGNCCTG 1980
DB 1886 CCGCGGAGTGGGCACTCTGCTCAACCGGCAAGTCCGAGAGACTTCTGTCGCGCGGT 1945
QY 1981 CGTGTGTGTGCGCCGCGCTCNCGTACTCTGTGCGCGGACGCGGCTCTGTACTGTTCG 2040
DB 1946 CGACTGTGTGCGCGGCTCTGTACTCTGTGCGCGGACTCAGATCTCTACGCTTCG 2005
QY 2041 TCGCGAGAGAGCTCGGCTTCAAGGCGCGCGGAGAGTCTTCTCGGCAAGGAGAGG 2100
DB 2006 TCGCGAGAGAGCTTGGGCTCAAGGCGCGCGGAGAGCTTCTCTCGGCAAGGAGAGG 2065
QY 2101 TGACGATCGGCAACAAGCTCTCCGATCTACAGAGGCATCAAGNCGGNGCATCAACC 2160
DB 2066 TGACGATCGGCTGGAAGCTCTCCAGATCTACAGAGGCATCAAGTGGGCAAGATCAACA 2125
QY 2161 ACGTCTCTGTGAAGATGCTGCNTAG 2186
DB 2126 ACGTCTCTCTCAAGATGCTGCNTAG 2151

Search completed: September 11, 2004, 14:50:34
Job time : 1376.94 secs

Db 241 TCGGCGCCGNCGCAAGGCGCCGNCNGTCGCGCTCNCAGACAGNCGACAGATCCGCGCA 300
Qy 301 AAGATCGACAAANAGNGTCGAGTTCTCTCCGNNCAGCTCNCAGACAGNGTCAAGGCTC 360
Db 301 AAGATCGACAAANAGNGTCGAGTTCTCTCCGNNCAGCTCNCAGACAGNGTCAAGGCTC 360
Qy 361 ACGATGTTTTCGCGGCTTCGCGGACACCCGAGATGAGATGCAATCTCGTCCAGAA 420
Db 361 ACGATGTTTTCGCGGCTTCGCGGACACCCGAGATGAGATGCAATCTCGTCCAGAA 420
Qy 421 GCGTCTTCGAGACACGATCTCGGCTGTCGACGTCGACGTCGACGTCGTCGTCGTC 480
Db 421 GCGTCTTCGAGACACGATCTCGGCTGTCGACGTCGACGTCGACGTCGTCGTCGTC 480
Qy 481 GCGGCGGCTTCGAGAACTGCTTCGCTCGAGTCTCGCGGCGCATGACATCCGC 540
Db 481 GCGGCGGCTTCGAGAACTGCTTCGCTCGAGTCTCGCGGCGCATGACATCCGC 540
Qy 541 GTCAACTGCTCAACGCGGCGCATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 541 GTCAACTGCTCAACGCGGCGCATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 601 AACTTCTCAACGCGGCGCATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 601 AACTTCTCAACGCGGCGCATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy 661 GCGGCGCTTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 661 GCGGCGCTTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy 721 GTNACGCTNNTNACGAGGCGCANNGAGAGATCATGTCGCGCGGAGCGATCGCGCTC 780
Db 721 GTNACGCTNNTNACGAGGCGCANNGAGAGATCATGTCGCGCGGAGCGATCGCGCTC 780
Qy 781 TTNGGCTTCGAGCGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 840
Db 781 TTNGGCTTCGAGCGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 840
Qy 841 GCGGCTTCGCGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 900
Db 841 GCGGCTTCGCGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 900
Qy 901 TCGAGGCGCTCAAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 960
Db 901 TCGAGGCGCTCAAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 960
Qy 961 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1020
Db 961 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1020
Qy 1021 TCGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1080
Db 1021 TCGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1080
Qy 1081 GAGAGGAGGAGGATTTCTCGCGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1081 GAGAGGAGGAGGATTTCTCGCGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Qy 1141 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1200
Db 1141 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1200
Qy 1201 TCGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1260
Db 1201 TCGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1260
Qy 1261 TTGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1320
Db 1261 TTGAGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1320
Qy 1321 GCGAGGCTCAACTTCAGGAGCTCAAGGAGTGTCAAGCGCGGAGTAACCGCGGCTT 1380
Db 1321 GCGAGGCTCAACTTCAGGAGCTCAAGGAGTGTCAAGCGCGGAGTAACCGCGGCTT 1380

Qy 1381 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1381 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Qy 1441 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1500
Db 1441 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1500
Qy 1501 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1560
Db 1501 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1560
Qy 1561 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1620
Db 1561 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1620
Qy 1621 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1680
Db 1621 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1680
Qy 1681 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1740
Db 1681 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1740
Qy 1741 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1800
Db 1741 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1800
Qy 1801 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1860
Db 1801 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1860
Qy 1861 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1920
Db 1861 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1920
Qy 1921 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1980
Db 1921 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 1980
Qy 1981 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2040
Db 1981 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2040
Qy 2041 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2100
Db 2041 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2100
Qy 2101 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2160
Db 2101 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2160
Qy 2161 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2220
Db 2161 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2220
Qy 2221 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2280
Db 2221 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2280
Qy 2281 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2340
Db 2281 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2340
Qy 2341 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2400
Db 2341 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2400
Qy 2401 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2460
Db 2401 GCGGCGCTTCGAGGCGCTGCTGCTGCGCGGAGAGGCTTGGTCTGCTCAAGCGACG 2460

Db	1717	CTGTGTGAGAGAGGTGAACAAACGGTCTGCCCAAGCGCCTTCAGAGACAACTCTGTACAC	1776
Qy	1801	CTTGAGACCCGCTGGGACGAGCGCTTCTTCGTTGGCGAACCGGACACCGTGTGTGAGNNNNN	1860
Db	1777	CTGTGTCCCGCTGGGACAGACGGCTTCTTCCTTCCGCGCGGACCCGTGTGTGAGGTCTCTC	1836
Qy	1861	NNGTCTCCGCGNNNNNGCCANNAGGTCTGCTCGGCGGCGTCAACGGCTGGAAGTCTGCT	1920
Db	1837	TCGTGACGCTC-----GCTCTGGCTCGCGCGCGTCAACGGCTGGAAGTCTGCGG	1885
Qy	1921	CGCGCGAGAGGCGCATCTTCGCTCAGCGGCGMANGTCTCGCGACGCCCTTCTGAGCGGCGCT	1980
Db	1886	CGCGCGAGGTGGGCACTCTGCGTCAACCGCGCAAGTCTCGGAGACTCTTCTGTCTCGCGGCT	1945
Qy	1981	CGTGTGCTGTGGCGCGGCTCNCGTACCTCTTGCGCGGACGCGCGTCTGTACTGTTCG	2040
Db	1946	CGACCTCGTGTGCGCGGCTCTGTACTCTGTGCGCGGCACTCAGATCTCTACGCTTCG	2005
Qy	2041	TCCGCGAGAGGCTCGGCGGTCAAGCGCGCGCGCGCGGCGGCGTCTCTCGGCGAAGCGAGG	2100
Db	2006	TCCGCGAGAGGCTTGGCGTCTTCAAGGCGCGCGCGGAGAGCTTCTCTCGCGAAGCGAGG	2065
Qy	2101	TGACGATCGGCAACCAACGTCCTCCGCACTTACGAGGCGATCAAGNNCGGCGNCATCAAC	2160
Db	2066	TGACGATCGGCTCGAAGCGTCTCAAGAGTCTACGAGGCGATCAAGTCTCGGCGAGATCAACA	2125
Qy	2161	ACGTCTCTGTCAAGATGCTCGGCTGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	2220
Db	2126	ACGTCTCTCTCAAGAGTCTCGGCTTACGACACTTTCGACACTCTTCCACTCTTGCAATCCCTTCAATCC	2185
Qy	2221	NN	2280
Db	2186	TATCCGCGCTGACACTCTTAGAGACTGTGCTTCTTGCGGACTCGGATCTCGCATCGCTCTT	2245
Qy	2281	TNN	2340
Db	2246	TCGTTCTTGGCTTGGCTCTTC-TAGACCGTTCGTTACTTGAAGTTGTGAATCAAGC	2304
Qy	2341	NGTNN	2400
Db	2305	AGTACCATTCACGACATCGATTAATCAGGAGAGAACTTACGCTTGGCGGAGCTTCTTG	2364
Qy	2401	NNNNNNNA 2408	
Db	2365	CGCATMAA 2372	

RESULT 3

US-09-624-693A-12

Sequence 12, Application US/09624693A

Patent No. 6355468

GENERAL INFORMATION:

APPLICANT: Yoshida, Roberta

APPLICANT: Koester, Anna

TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and

TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and

TITLE OF INVENTION: Using Same

FILE REFERENCE: 29479/500NSC

CURRENT APPLICATION NUMBER: US/09/624,693A

CURRENT FILING DATE: 2000-07-24

NUMBER OF SEQ. ID NOS: 25

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 12

LENGTH: 2419

TYPE: DNA

ORGANISM: Rhodocorulia graminis

FEATURE:

NAME/KEY: CDS

LOCATION: (37)..(2196)

US-09-624-693A-12

Query Match 65.8%; Score 1629.4; DB 4; Length 2419;

[illegible]

Db 1946 GGAACCTGCGCGCGCTGTAACCTTCGCCGCGACACAGATCCTTAAGCCTTCG 2005
QY 2041 TCCGGAGAGAGTCCGGCTCAAGGCGCGCGCGCGCGCGCTTCCTCGGCAACAGAGG 2100
Db 2006 TCCGGAGAGAGTTCGGCTCAAGGCGCGCGCGCGCGCGCTTCCTCGGCAACAGAGG 2065
QY 2101 TGAAGATCGGACACAGCTTCGCCGATCTACGAGGCATCAAGNNCGGNGATCAACC 2160
Db 2066 TGAAGATCGGCTGGAAGCTTCGAAGATCTACGAGGCATCAAGTCCGAGATCAACA 2125
QY 2161 AGCTCTCGTCAAGTCTCGCTG 2186
Db 2126 AGCTCTCGTCAAGTCTCGCTG 2151

RESULT 6
US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sarielani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BCI009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodocorula glutinis
US-09-627-216A-7

Query Match 65.6%; Score 1623.2; DB 4; Length 2151;
Best Local Similarity 83.6%; Pred. No. 3.6e-291;
Matches 1827; Conservative 0; Mismatches 324; Indels 35; Gaps 5;

QY 1 ATGGCCCTCCCTCCGATCTGATCGGACCTCGATCGGCAAGGCTGCAAGGCTG 60
Db 1 ATGGCCCTCCCTCCGATCTGATCGGACCTCGATCGGCAAGGCTG 60
QY 61 CAGCGCCGNNCGCANNNGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 61 CAGCGCTG-----CAATGGGCGCTGACCACTC-----GCAGTCGAG 100
QY 121 GCTCGCTCTCCCGACCAACCCAGGAGAGTGCAGATCGTGNAGNAGATCCCTGCGG 180
Db 101 GCTCGCACTCGCCCAACCCAGGAGTGCAGATCGTGNAGNAGATCGTGC 160
QY 181 ACCCCACCGNACGNAAGNNNTGAACTCGAGCGGATCAACCTCTCAACCTCGGNGAG 240
Db 161 CGCGGACCGACTGA---CGCTGAACTCGAGGCGTCACTCGTCAACCTCGGAGAGT 217
QY 241 TCGGCGCGCGCGCAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 218 TCTGCGCGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
QY 301 AAGATGACANAGAGTCTGAGTCTCTCGGNNCGAGCTGACAGACAGAGTCTACGAG 360
Db 277 AAGATGACAAATGCTGAGTCTCTCGGCGGCGGCGGCGGCGGCGGCGG 336
QY 361 ACGATGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 337 ACGATGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
QY 421 GCGCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

Db 397 GCTCTCTCGAGGACCAAGCTCTGCGGCTTCCTTCCTTCGATCGAGCTGTTCCGCGCTC 456
QY 481 GCGCGGCGCTCGAGAACTGCTTCGCTCGAGGCTCTCGGCGGCGGCGGCGGCGG 540
Db 457 GCGCGGCGCTCGAGAACTGCTTCCTTCGAGGCTTCGCGGCGGCGGCGGCGGCGG 516
QY 541 GTCAACTCGTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 517 GTCAACAGCTTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
QY 601 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 577 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
QY 661 GCGGAGCTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 637 GCGGAGCT 696
QY 721 GTCACGCTNNTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 697 GTGACGCTGTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
QY 781 TTGAGTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Db 757 TTCAACCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
QY 841 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Db 817 GCGGCTGAGATGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 876
QY 901 TCGAGGCGGCTGACGCGGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Db 877 TCGAGGCGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
QY 961 CCNTTCTTCAGAGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 937 CCNTTCTTCAGAGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 996
QY 1021 CGCAGCTCTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 997 CGCAGCTCTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
QY 1081 GAGCAGAGGAGATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Db 1057 GAGCAGAGGAGATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
QY 1141 GCGCGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Db 1117 GCGCGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
QY 1201 TCGAGCAGCGGCAACCCGCTCATGAGTGCAGAGCAAGNAGACCAACCGGCGGCGG 1260
Db 1177 TCGAGCAGCGGCAACCCGCTCATGAGTGCAGAGCAAGNAGATGCGGCGGCGG 1236
QY 1261 TTCCAGGCGGCGGCTGTCGNAACGATGAGAGAGCTCGCTCGGCTGCGGCTGATC 1320
Db 1237 TTCCAGGCGGCGGCTGTCGNAACGATGAGAGAGCTCGCTCGGCTGCGGCTGATC 1296
QY 1321 GCGAGCTCACTTACGAGGCTGACGAGTCAACGCGGCGGCGGCGGCGGCGGCGG 1380
Db 1297 GCGAGCTCACTTACGAGGCTGACGAGTCAACGCGGCGGCGGCGGCGGCGGCGG 1356
QY 1381 CCNTTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
Db 1357 CCNTTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
QY 1441 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
Db 1417 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
QY 1501 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
Db 1477 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536

1081 GACGACGAGGAGATTTCTCCGACGAGACCGCTACCCGCTCCGCAAGTGCCTCAGTGCCTC 1140
Db 1057 GACGACGAGGAGATTTCTCCGACGAGACCGCTACCCCTTGCGCAAGTGCCTCAGTGCCTC 1116
Qy 1141 GGGCCGCTCGTCAAGCAGATGATTCACGCGCCACGCGTCTCTGCTGAGGCGGACG 1200
Db 1117 GGGCCGCTCGTCAAGCAGATTCATTCACGCGCCACGCGTCTCTGCTGAGGCGGACG 1176
Qy 1201 TCGACGACGAGACACCCGCTCATTCAGCTCGAGAACGAGACGACGAGGAGGAGC 1260
Db 1177 TCGACGACGAGACACCCCTCTCATTCAGCTCGAGAACGAGACGAGGAGGAGG 1236
Qy 1261 TTCGAGGACGACGCTGCTCGACCAACGATGAGGAGGAGGAGGAGGAGGAGGAGG 1320
Db 1237 TTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296
Qy 1321 GGGAGGCTCACTTCACGAGCTCACCGAGATGCTTCAAGCGCGGAGGAGGAGGAGGAGG 1380
Db 1297 GGGAGGCTCACTTCACGAGCTCACCGAGATGCTTCAAGCGCGGAGGAGGAGGAGGAGG 1356
Qy 1381 CCNTCCGCTGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Db 1357 CCTCTGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1416
Qy 1441 GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db 1417 GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1476
Qy 1501 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Db 1477 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
Qy 1561 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
Db 1537 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
Qy 1621 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
Db 1597 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1656
Qy 1681 NCNCGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Db 1657 GTCGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1716
Qy 1741 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
Db 1717 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1776
Qy 1801 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
Db 1777 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1836
Qy 1861 NNCTGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
Db 1837 TCGTGCAGGCTC-----GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1885
Qy 1921 CCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
Db 1886 CCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1945
Qy 1981 CCGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
Db 1946 CCGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2005
Qy 2041 TCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
Db 2006 TCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2065
Qy 2101 TGAAGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
Db 2066 TGAAGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2125

Qy 2161 ACGTCTCTCTCAAGATGCTGCGCTAG 2186
Db 2126 ACGTCTCTCTCAAGATGCTGCGCTAG 2151

RESULT 8
US-09-765-873A-31
Sequence 31, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 31
LENGTH: 2061
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2061)
US-09-765-873A-31

Query Match 65.1%; Score 1610.4; DB 4; Length 2061;
Best Similarity 85.0%; Pred. No. 8.2e-289;
Matches 1764; Conservative 0; Mismatches 297; Indels 15; Gaps 3;

Qy 111 GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 170
Db 1 GCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
Qy 171 ATCTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230
Db 61 ATCTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 117
Qy 231 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290
Db 118 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 176
Qy 291 GATCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
Db 177 GATCGCGTCAAGATGAGCAATGAGTGAATGCTGAGTCTGCGTGCAGACTTCGATGAGCGT 236
Qy 351 CTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
Db 237 CTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 296
Qy 411 GCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 470
Db 297 GCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 356
Qy 471 CTTGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 530
Db 357 GTTCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 416
Qy 531 GACCATCGCGTCACTGCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 590
Db 417 GACCATCGCGTCACTGCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476
Qy 591 GGGGCTCACCACCTTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 650
Db 477 GGGGCTCACCACCTTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536
Qy 651 CTGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 710
Db 537 CTGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 596

```

QY 711 GGAACNNCAAGTNCACGNTNTNACAGAGGGCAANNAGATCATGTNCGCCGGGAGGC 770
Db 597 GGAACAGAAAGTGCACGTGCTCCACAGAGGGCAAGAGAAAGTCTGTAGCCCGGAGGAC 656
QY 771 GATCGCGCTCTTNGTGTCTGAGGCCGTGCTCTCGGCCGGAAGAGAGGTCTCGTCTGCT 830
Db 657 GATGCGCTCTTCAACCTTCAGCCCTGAGCCGCTCGGCCGGAAGAGAGGTCTCGTCTGCT 716
QY 831 CAACGGAGAGCGCGTCTCGCGCTCGATGCGGACCCCTCGCTCTGACGACGACATATGCT 890
Db 717 CAAGGAGACCGCGCTCTGACAGATGATGACACCTCGCTCTGACGACGACATATGCT 776
QY 891 CTGCTCTCTCTGACAGCGCTCTGACGCTCTGACGCTCTGACGAGGAGGAGTCTGAGTCTG 950
Db 777 CTGCTCTCTCTGACAGCGCTCTGACGCTCTGACGCTCTGACGAGGAGGAGTCTGAGTCTG 836
QY 951 CTGCTCTCTCTGACAGCGCTCTGACGCTCTGACGCTCTGACGAGGAGGAGTCTGAGTCTG 1010
Db 837 CTGCTCTCTCTGACAGCGCTCTGACGCTCTGACGCTCTGACGAGGAGGAGTCTGAGTCTG 896
QY 1011 GCGCAACATCCGACGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1070
Db 897 GGGAAACATCCGACGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 956
QY 1071 CAAGGTCAAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1130
Db 957 CAAGGTCAAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1016
QY 1131 TCAGTGGCTCGGCGCGCTCTCTGACGAGCATGATGATGATGATGATGATGATGATGATGAT 1180
Db 1017 TCAGTGGCTCGGCGCGCTCTCTGACGAGCATGATGATGATGATGATGATGATGATGATGAT 1076
QY 1191 GGCCTGACGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1250
Db 1077 GGCCTGACGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1136
QY 1251 CGCGCGCACTTCCAGGCGCGCTCTGACGAGCATGATGATGATGATGATGATGATGATGATGAT 1310
Db 1137 CGCGCGCACTTCCAGGCGCGCTCTGACGAGCATGATGATGATGATGATGATGATGATGATGAT 1196
QY 1311 CGCGCTGATCGGAGAGTCACTTCAAGCAGTCAAGCAGTCAAGCAGTCAAGCAGTCAAGCAGT 1370
Db 1197 CGCGCTGATCGGAGAGTCACTTCAAGCAGTCAAGCAGTCAAGCAGTCAAGCAGTCAAGCAGT 1256
QY 1371 CGCGCGCTTCCGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1430
Db 1257 CGCGCGCTTCCGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1316
QY 1431 CCTCGACATTTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1490
Db 1317 CCTCGACATTTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1376
QY 1491 GACCCAGCTTCAGCGCGGAGAGATGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCT 1550
Db 1377 GACCCAGCTTCAGCGCGGAGAGATGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCT 1436
QY 1551 GGCNCGCGCAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1610
Db 1437 GGCNCGCGCAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1496
QY 1611 CTGCGCTTCTGAGCGCGTGCAGCTCGCGCGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1670
Db 1497 CTGCGCTTCTGAGCGCGTGCAGCTCGCGCGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1556
QY 1671 CCCGATNTTCCNNGGCTCTNTAGAGAGCATTTTGGGCGGCGCTCTGAGCGGCGGCGGCGGCGG 1730
Db 1557 CCCGATNTTCCNNGGCTCTNTAGAGAGCATTTTGGGCGGCGCTCTGAGCGGCGGCGGCGGCGG 1616
QY 1731 NNNNNAGAGTTCAGGAGCAGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGTCAAGAGGAGT 1790
Db 1617 GCGGAGCGAGGCTCTGTCAGAGAGTGAACAAGAGTCTGCGCAAGGCGCTCTGAGGAGAGACCA 1676

```

```

QY 1791 CTGTAAGACCTCGAGCGCGGCTGCGACAGACGCTCTCTGTTGCGGAGCCGAGACCGTCTG 1850
Db 1677 CTGTAAGACCTGCTCCGCGCTGCGACAGACGCTCTCTCTGCGCGCGAGCCGCTGCT 1736
QY 1851 CGAANNNNNNNGTCTCTCGCCNNNNNCCANNAGAGTCTGCTGCGGAGGAGTCAAGCGCTG 1910
Db 1737 CGAGTCTCTGTCGAGAGTCT-----GCTCTGCTGCGCGGAGTCAAGCGCTG 1785
QY 1911 AAGTGGCTCTCGCGGAGAGGAGCAGTCTGCTCAAGCGCGGAGGAGGAGGAGGAGGAGGAGT 1970
Db 1786 AAGTGGCGCGCGCGAGTGGCGCATCTGCTCAACCGCGCAAGTCCGAGAACCTTCTG 1845
QY 1971 NCGGCNCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
Db 1846 TCCGCGCGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1905
QY 2031 TACTGCTTCTGCTGCGGAGAGTCTGCGCTCAAGGCGCGCGCGGAGAGTCTTCTGCTG 2090
Db 1906 TACGCTTCTGCTGCGGAGAGTCTGCGCTCAAGGCGCGCGCGGAGAGTCTTCTGCTG 1965
QY 2091 AAGCAGAGGAGTACGATCGGAGCAGACGCTCCGAGCTCAAGGAGGAGGAGGAGGAGGAGGAGT 2150
Db 1966 AAGCAGAGGAGTACGATCGGAGCAGACGCTCCGAGCTCAAGAGTCTCAAGGAGGAGGAGT 2025
QY 2151 NCATCAACACGCTCTGCTCAAGATGCTCGCTAG 2186
Db 2026 AGGATCAACACGCTCTCTCAAGATGCTCGCTAG 2061

RESULT 9
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 64.6%; Score 1598.6; DB 4; Length 2787;
Best Local Similarity 82.9%; Pred. No. 1.3e-286;
Matches 1812; Conservative 0; Mismatches 330; Indels 44; Gaps 5;

```


NUMBER OF SEQ ID NOS: 25
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 14
 LENGTH: 2311
 TYPE: DNA
 ORGANISM: *Amarilia muscaria*
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (18)..(2237)
 US-09-624-693A-14

Query Match 13.0%; Score 322.6; DB 4; Length 2311;
 Best Local Similarity 52.8%; Pred. No. 3e-51;
 Matches 734; Conservative 0; Mismatches 625; Indels 30; Gaps 4;

```

QY 345 CAGNGCTTACGAGTACGACTGCTTTGCGCGGCTTCGCGACACCCGAGTGAAGATGC 404
DB 368 CAGGATATACGAGCTGTCACTGTTTCCGTGGCAGTGTCTGATACAGGACGACCAACC 427
QY 405 NATCTGCTCCAAAGGNCCTCTCGAGACACAGCTGCGGTGTCCTCCGACCTGCT 464
DB 428 GATGTTTGGGGTTTGGCTTTTGGCAACCAACTGAGGATATCTGCCACCTCGAC 487
QY 465 CGATCTTCGCTCGAGCGCGGCTCGAG-----AACTGCTTCGCTCGAGGT 515
DB 488 TGAACCTTTGAGCTCTACTCTCTCAAGATGCAATAACAGACATGCCAGAGCGTG 547
QY 516 CGTCCGGGGGCGCATGACATCCGCGCTCACTCGCTCANCGCGGCGCACTGGCNGTCCG 575
DB 548 GATTCGGGGGGCATTTTATCCGATATGAAATTCGCTGATTCGCGCACTCGAATAG 607
QY 576 CCTGCTGCTCTCGAGCGCTCCCAACTTCTCTCAACAGGACATCAACCCCATGCTCC 635
DB 608 ATGGAGTTGATGAAAAGATGAGAACTACTCGCGGCAATGATACCTGTCTCC 667
QY 636 CTTCCGGGCACTATTCGGCGTGGGCGAATCTTCCGCTCTCTAATATGCGCGCGC 695
DB 668 CCGAAGAGGAGCATCTCTCACTCGAGATCTGCTCCCTATCTATTCGAGGAC 727
QY 696 CATCACCGGTCAACCGGACNNCAAGTNCAGCTNNMNC-----ACGAGGACNNAGAA 749
DB 728 GATTATTTGGCAACCCATCAATCAAGTATATACAGCTCATTAAGTCCGAAATTCGCCA 787
QY 750 GATCATGTCGCGCGGAGGAGCATGCGCTCTTNGATCTCGAGCCGCTGCTCGCGCC 809
DB 788 AATTGATCTCTGAAGATGTCTTGCTGTGATATATGAACTTTCCACATGGAATC 847
QY 810 GAAGAGGGTCTCGCTGCTGCTCAACGCGACGGCGCTTCGCTGATGGGACCCCTGCC 869
DB 848 GAAAGAACTCTTGTATTTTAATGGACCGCATCTCGGATCTGGGAGCTTTAGC 907
QY 870 TCTGACGACGACACATGCTCTGCTCTCGAGGCGCTCACGCTCTNACGCTGCA 929
DB 908 CCTAAACGAAAGCATCATCTTGTCTGTGCTCAAGTGTGACGCTATAGGGACCGA 967
QY 930 GGGCATGCTGCGCGACGCGGCTGCTTCCACCCNTTCTTCCACGAGTACGCGCCCTCA 989
DB 968 GGGATGATGAGGACTCGGCTTCTCATGACCGCTTCAATTCATGCGACCGACACACA 1027
QY 990 CCCGACCCAGATGAGGTGCGCGGCAACATCCGCAAGCTCCCTCGAGGGGACGNNGTTGC 1049
DB 1028 TCCCGGTCAAGTATGATGTCTGAGAACTTTGGAATTTGCTCGATGGGAGTAATTGGC 1087
QY 1050 CGTCCACCAAGAGAGAGGTAAAGTCAAGACGACGAGGACATTTCTCGCCACGACCG 1109
DB 1088 TCAGTTAGAAAGACAGAGTTGCTTAGAAGACGATTAATACACCTTCGCGAGACCG 1147
QY 1110 CTACCGGCTCCGAGGTGCGCTCAAGGCTCGCGCGCTCGTCAAGGACATGATTCACGC 1169
DB 1148 TTTTCCACTCCGAATCTTCCCTCAATCTTGGGCTTGAATGACATTAATCTCCGC 1207
QY 1170 CCAAGCNGTCTCTGCTCGAGCGCGACAGT---CGACGACCGCAACCCGCTCATGCA 1226

```

```

DB 1208 TTTCCAGACTGTAAACGAGAGTGTATTAATTACTTAACAGTACTGCAATTCAGTATTGA 1267
QY 1227 CGTGAAGAACAGANGAACCCACACGCGCGCAACTTCCAGAGCGGCTGGCNAACAC 1266
DB 1268 TGGGAGACTGGGGAATCTCACACGATGCGCAATTTCCAGAGATGGCTGTACTAATCC 1327
QY 1287 GATGAGAAAGACTCGCTTCGNCCTGCGCTGTATCGGCAAGCTCAACTTCAGCAGCTCAC 1346
DB 1328 AATGAGAAAGACGCGACTTGTCTTAATCACTGTCGAATTAATTTCCAGAGAC 1387
QY 1347 CGAGATGCTCAACGCGGCGATGACCGGCGCTTNCCTGCTCGCTGCGAGGAGACC 1406
DB 1388 TGATTTAGTCAATCTTGCATTAACCGGCTTTCGCGCTTCACTGCTGCGACGATCC 1447
QY 1407 NTGCTCTCTTATCACTGCAAGGCTCGACATTCGCGCNGCNGTACATTCGAGACT 1466
DB 1448 ATCTCACTACAGCGCCAAAGACTAGACATAGCACTGGGCTTACGAGCCGA--- 1504
QY 1467 CGGNCACCTTGGCAACCCGTTAAGACCCAGCTCCAGCGCGCNGAGATGGGCAACAGGC 1526
DB 1505 -----AGGACTCTGGCGCCCACTCAATTAATGTCGAGAAATGCAACCAAGC 1555
QY 1527 CGTCAACTGCTCGAGCTCATCTCGGCGCAGCGCACGCGCAAGGCAACGACGTCTTC 1586
DB 1556 TGTTAATCTCCGTGGGTGATTTCTGCTCGGCTACATCAATGCTTGGAGAGTCTAAC 1615
QY 1587 TCTCTCTCTCGCACCCACTTACTGCGTCTCCAGGCGCTCGAATCTCGCGGATGGA 1646
DB 1616 ATCTGATGCGGCTTACTGTATATTTCTATGCCAGACTCTGACCTCCGCTTCA 1675
QY 1647 GTTCAGTTCAAGAAAGAGTTCGACCCGNTNNTCNNCGCTCNAAGACGACTTTGG 1706
DB 1676 GCGGAGTTCTTGGCGCGGTCTAGACATCATCTGCTGAGAGTTAAGATGCTATTGG 1735
QY 1707 CNGCGCTT 1715
DB 1736 ATCTTCTT 1744

```

RESULT 11
 US-09-615-192A-245
 ; Sequence 245, Application US/09615192A
 ; Patent No. 6410718
 ; GENERAL INFORMATION:
 ; APPLICANT: Bloksberg, Leonard N.
 ; APPLICANT: Havukela, Ilkka
 ; TITLE OF INVENTION: Materials and Methods for the
 ; TITLE OF INVENTION: Modification of Plant Lignin Content
 ; FILE REFERENCE: 11000.1003c4U
 ; CURRENT APPLICATION NUMBER: US/09/615.192A
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 06/975,316
 ; PRIOR FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: US 08/713,000
 ; PRIOR FILING DATE: 1996-09-11
 ; PRIOR APPLICATION NUMBER: US 09/169,789
 ; PRIOR FILING DATE: 1998-10-09
 ; NUMBER OF SEQ ID NOS: 405
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 245
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: *Eucalyptus grandis*
 ; US-09-615-192A-245

Query Match 8.6%; Score 213.2; DB 4; Length 1455;
 Best Local Similarity 54.8%; Pred. No. 4.4e-31;
 Matches 480; Conservative 0; Mismatches 384; Indels 12; Gaps 4;
 QY 465 CGATCTTCGCTCGAGCGCGCTCGAGACTGCTTCGCTCGAGTGTGCGCG 524
 DB 592 CCGATCTTCGCAACGACGAGAGTGTGCGCACACCGCTGCTCAATCTCCACCGAGC 651

Mon Sep 13 10:31:13 2004

us-09-939-408a-20.rnd

Page 14

```

QY      522  CGGCATGACCATCCGGGTAACTCCGGTACACMGCGGCACATCCGACGCTCGTGCAT  584
Dn      652  CGGCATGCTGTCCGGGTAAACACCTCTCTCCAGGGCTACTCCGATCGTTTGAGAT  711
QY      585  CCTCGAGGCGCTCACCACTTCTCTCAACCCAGGCTACACCCCATGTCCTCCCGCG  644
Dn      712  CCTCGAGGCGCATCACCAAGTTCCTCAACCAACATCACCCCGGTGCTCGCCCTCA  771
QY      645  CACCATCTCGGCGCTCGGGGAGACCTCTCCCGACCTCAACATCGCGCGCGCATCAC  704
Dn      772  CACCATCATCGCTTCAGCGCATGTGTCTCCCTCTCTCATATTGCGGGGCTTCGAG  831
QY      705  TCACCCGGAACNNCAAGTTCACGTNNNTNACAGAGGGCANNAGAAAGATCATGTC  764
Dn      832  CCGGCCCAATCCAAAG---CGTGGGGCTGATGGAAGTCCCTGACGCTGTCGAGGC  888
QY      765  CGAGGCGATGCGGCTTTNAGTCTGAGGCCCTGCTCTCGGCGCCAGAGAGGATCTCG  824
Dn      889  CTTCGGCTTCGCGGATGGAACAAGGGCTTCTTTCAGCTGACGCCCAAGAGAAAGG  948
QY      825  TCTCGTCAACGACGACGCGCTCTCCGCTTCAGTGCAGACCTCGCTTCGACGACGA  884
Dn      949  GCTGTGAAAGGACGACGCGCATCGGGCTGCGCTGCTTCGATCGTCTTTCGAGGCA  1000
QY      885  CATGCTCTGCTCTCTCTCGAGGCGCTCAAGGCTCTAAGTTCGAGGCTCATGTGCG  944
Dn      1009  CATACTGCGGCTCTCTCGAGGCTCTCTCAACGATCTTCGAGAGGATGACAGGGAA  1060
QY      945  CGCGGCGTGTCCACCCNMTCTCTCAGACGTCACGCGGCTTCCCGACCCGATTCGA  1000
Dn      1069  GCGGAGTTTACAGACCACTTACGATTAATTGAGACCACTCCGG---CAGATTGA  1122
QY      1005  GGTGCGCGCACATCCGACGCTCTCTCGAGGCGCANNNGTTTCCGTTCACACGAGA  1060
Dn      1126  GTCTGGGCTATATAGAGACATTTTGGATGAAGCGC--TTCAGTGAAGGCTGCTAA  1184
QY      1065  GGAAGTCAAGTCAAGACGACGAGGGAATTTCCGACAGACCGCTACCCGCTCCGAC  1122
Dn      1183  AAGATTGACAGAGATGATTCGCTCCAGAAAGCCAAAGACAGAGTACGTTCTTAGAC  1220
QY      1125  GTGCGCTCACTGCTCGGCGCGCTGCTCAGCACAATTCAGCGCCACGCGTCTCTC  1184
Dn      1243  TTCTCCCAATGGCTAGGGC---CCAGTTGAGTGATCCGAGCGGACCAAGATAT  1220
QY      1185  GCTCGAGGCGCGACGTGACGACCGACACACCCGCTCATCGACGTGAGAAACAGAAC  1220
Dn      1300  TGAAGGGAATCATATTGCTCATATACACCCGCTGATCGATGTGCGAGAAACAAGC  1330
QY      1245  CCAACCAAGGGGCGCACTTCAGGCGACGCGCTGTGCGMAACAGATGAGAAAGTCCCT  1330
Dn      1360  CTTGCAAGGAGGAACTTCAGGAGACCCCGATTGTTGCTCATGACACACTCGCT  1440
QY      1305  CGNCCTGCGCCTGATCGGCAAGCTCACTTCAGCA  1340
Dn      1420  GCGGTTGCTCATATAGGCAAGCTCATGTTGCGCA  1455

RESULT 12
US-09-615-192A-247
; Sequence 247, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukala, Ilka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.100354U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	6.9%;	Score 170.2;	DB 4;	Length 1520;
Best Local Similarity	49.7%;	Pred. No. 3.8e-23;		
Matches 480;	Conservative	0;	Mismatches 473;	Indels 13; Gaps 4

QY	791	AGCCCGCTGCTCGGCCCCAAGGAGGATTCGATCTGTCAACGCAACGCGCTGCTCG	851
Db	124	AGCCCTTGAATGAGGCTTAAGAAAGTGTGGCCATTGTCAATGGCACTTCAGTGGAG	183
QY	851	CCTCATGCGCAACCTCGCTCTGCAAGCAACATGCTCTGCTCCTCTTGCAGCGGC	910
Db	184	CAGCACTGGCTTCCATTGTGTGTTCATGCGCATGTTCGTCTGCTCTCTGAAATPA	243
QY	911	TCAAGGCTTMAAGGTGAGGCGCATGTCGGGCAACCGCGACTGATTCACCCNTTCTCC	970
Db	244	TCTCTGCATGTTCTGCGAGGTTATGTAATGGAAAGCTGATTTACATTCATTAACTC	303
QY	971	ACGAGCTCACGCGCCCTCACCGAACCCGATGAGGTGCGCGCAACATCCGACGCTCC	1030
Db	304	ACAGGTGAAGACCAT---CCTGGCCAATGGAACCTCAGCGATCATGAGTATGTCT	360
QY	1031	TGAGGCGAGCNGNTTTCGCTCCACACGAGGAGGATCAAGATCAAGAGCAACGAGG	1090
Db	361	TGAGCGGGA---GCTTATATGAAACGCTCTTAAGCTCATGATGAATCTCTGCG	418
QY	1091	GCATTCCTCGGCGAGGACCGGTACCCGCTCGGCAAGTCGCTCAATGAGTCTCGGCGCTCG	1150
Db	417	AGAAAGCAAGAGGATGCTATGCGGTTTGCACTTGCTCAGTGGCTCGGCGCTCAGG	476
QY	1151	TCAGGCATGATTCACGCCCAACGCGNCTCTCTGCTCCAGGCGGAGTCAGTGAACCG	1210
Db	477	T---GGAGATTATCAGATCTGCACTCACATGATTTAGCGGGAATCAATTGTGTAATG	533
QY	1211	ACAAACCGCTCATGACGCTGAGAAACAAGANGAACCAACGACGCGCAACTTCCAGCGGN	1270
Db	534	ACAAATCAGTAATTAATGTTGCCAGAACAAAGCTTCAATGAGAGGAAATTTCCAGGCA	593
QY	1271	CCGCTGTGCNAACGATGAGAAAGCTCGCTCGGCTCGGCTCGCTGATCGGCAAGCTCA	1330
Db	594	CACCATATGATGTTCCATGATATATCTTGCTGTGCAATTCAGCAATTTGGGAATTTGA	653
QY	1331	ACTTCAACGAGTCAACGAGATGCTCAACCGCGGATGAACCGCGGCGCTTNCNTCTCGC	1390
Db	654	TGTTGCTCAATTCACAGCTTGTGATGATTTACTACAAATGAGAGCTTGCTTGGAAATC	713
QY	1391	TGCGTCCCGAG---GACCNTCGCTCTCTCATCACTGCAAGGCGCTGCACTTGC CGNG	1447
Db	714	TGATGTGTGGGCTAATCCAGCTGATTAATGACATGAAAGGGCGGAGATCGCTATG	773
QY	1448	CNCGNTACCTTGGAGTCCGAGNACCTTGCCAAACCGGTMACGACCGCGTCCGACCGG	1507
Db	774	CTTCTTACACTTCTGAGCTTCTTTACTCTGGGAATCTGTGCACAGCCATGTACAGAGCG	833
QY	1508	CNAGATGAGGCAACGAGCGCTCAACTCGCTCGGCTCACTTCGCGNCCCGCACNCGCG	1567
Db	834	CCGAACAGATTAACAGGATGTCAATTCCTGGGTCGTTCAAGTAGAAATCTGCGCG	893
QY	1568	AGGCCAAAGACGTCTTTCTCTCTCTCTCGCACCCACCTCTACTGCGGNTCTCAGCGCG	1627
Db	894	AGGCCATCATATTCCTGAAGCTGATGTCTTCCACATACCTGACAGCTCTGTGCGCAGGCTG	953
QY	1628	TCGACCTCGGCGAGTGAAGTTGAGTTCAAGAAAGAGTTCGACCGCGGANTTNCNNCGC	1687
Db	954	TGATTTAAGGATCTGAGGAAAAAATGCTGGCGCACTGTAAGCAGATTTGTTCTCAGG	1011

QY 1688 TCNTCAGACGACGACTTTGGCNCNGCCCTCGACGCGNNNNNNNNNNNACGAACTCGANG 1747
 DB 1014 TAGCCAGAAACCTCGACGACGAGGCTCAAGGAGGAGCTTTTGCAGCGCGTTTTCGCG 1073
 QY 1748 ACAAGG 1753
 DB 1074 AAAAGG 1079

RESULT 13

US-09-252-991A-16475
 ; Sequence 16475, Application US/09252991A
 ; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16475

; LENGTH: 1548

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16475

Query Match 5.3%; Score 131.6; DB 4; Length 1548;

Best Local Similarity 46.5%; Pred. No. 5e-16; Mismatches 606; Indels 52; Gaps 5;

Matches 572; Conservative 0; Mismatches 606; Indels 52; Gaps 5;

QY 471 CTTNGCCTCGGCGGCGCTCGAGACTCGCTCGAGTCTCGCGCGCGGCAT 530
 DB 336 CTTGCTCCATGCGCGCGGCTCGGAGCGCTGAGACGACCAATGATGCGGTGTCAT 395
 QY 531 GACCATCGCGCTCACTGCTCAACGCGGCGCACTCGGCGGCTCGCTCGCTCGA 590
 DB 396 GTGCTCAAGGTAACAGCTGGCGGCTTCTCGGCAATCGCGCGCAAGGTGATGA 455
 QY 591 GCGGCTCAACCACTTCTCAACGAGCATCAACCCCATCTGCCCTCGCGGACCAT 650
 DB 456 CGCGCTGATCGCGCTGATCAAGCCGAGGTCTATCCGACATCCCGCTGAAGGCTGGT 515
 QY 651 CTGCGCTCGGCGGCGCTCTCCCNCTCTCTCAATCGCGCGGCATACCGGTCAACC 710
 DB 516 GCGGTGCTCGGCGGCGCTGCGCGCGCTGCGGACATGCTGCTGATCGGCGAAG 575
 QY 711 GGAACNNCAAGTNCAGTNTTCAAGAGGCAANNAGAAATCATGTCGCGCGCGAGGC 770
 DB 576 CCGGCGCGCGCATCGGGTG-----AGTGGCTCGCGGCGCGCGAAGC 617
 QY 771 GATCGGCTCTTNGTCTCGAGCCCGCTGCTCGCGCGCGAAGAGGCTCGGTCTCGT 830
 DB 618 GCTGGGCGGTGGCGGCGTGAAGCGCTGACCTGCGCGCAAGAGGCGCTGCGCTGCT 677
 QY 831 CAACGCGAAGCGGCTTCTCGCTGATGCGAAGCTCGCTCTGACGAGCAACATGCT 890
 DB 678 CAATGCGACCCAGGTCTCAACGCGCTTAAGCGGTGCGGAGTGTTCAGAGCGGAGACT 737
 QY 891 CTGCGCTCTCTGAGAGGCGCTCAACGCTCTNAAGTCTGAGGCGCAATGCTGCGCG 950
 DB 738 GTTCGCGCGCGCGCGCTGCGGCGGCTCAAGCGTCAAGGCGCATGCTCGTTTGGCGGG 797
 QY 951 CTGCTTCAACCCNTTCTCAACGAGTCAAGCGGCTTCAACCGACCAAGATCGAGTTCG 1010
 DB 798 GCGCTTCAATGCGCGCATCGCGCGCGGCGGCGGAGTGTGAGAGATCGACGTGCGCG 857
 QY 1011 GCGCAACATCCGAGCGCTCTCGAGGCGAGGNNGTTCGCTCCACCAACGAGAGGAGT 1070

DB 858 GGCCTATCGGACCTGCTCACCGCGAGGAGGTGCGGCTCCCATGAGAGTGCAG 917
 QY 1071 CAAGTCAAGACGACGAGGCAATTTCCGCGAGACCGCTACCCGCTCGGACGTGCC 1130
 DB 918 CAAGT-----CCAGAACCTTATTCGCTGCTGCCAGCC 953
 QY 1131 TCAGTGTCTGCGCGCGCTGTCAGCGACATGATTCACGCCCGAGGCTCTCTCGTCA 1190
 DB 954 GCAGGTGATGGCGCGCTGCTGACCCAGATGCGCCAGCGCGGAGGTGTGAGATCGA 1013
 QY 1191 GCGCGGAGTGCAGACCGCAACCGCTCATGAGCTGAGAGAACAGNAGCCCA 1250
 DB 1014 AGCC--AACGGGTGTCCGACAAACCGCTGTGTTTCGCGCGAGGCGACGTATTC 1070
 QY 1251 CCGCGGCACTTCCAGGCGGCGGTGTGCAACAGATGAGAGACTCGCTCGGCT 1310
 DB 1071 CCGGCGCACTTCCAGGCGGCGGCAACCGGTGAGATGCGCGGACACTGCGCTG 1130
 QY 1311 CCGCTGATCGGCAAGTCAACTTCAAGCACTCAACGAGATGCTCAAGCGCGGATGAA 1370
 DB 1131 GCGCGAGATCGGTTCGCTGTCGAAACCGGCACTGCTGATGAGACATGACATGTC 1190
 QY 1371 CCGCGGCTTCCGCTGCTGCTGCGAGGAGCCGCTGCTCTATCATGCAAGG 1430
 DB 1191 GCAGTGCGCGCGCTTCTCGGTGCGCAACGCG-----GGGTCACTCGGCTTCATGAT 1244
 QY 1431 CTTGCAATTCGCGCGGCGGCTGCTGAGTCTGAGCTGAGACCTTCCAAACCGGTTAC 1490
 DB 1245 CCGCGGATGCAACCGCGCGCGCTGCGGCGGCAACAAAGGCTTGCCTGCGGCGAG 1304
 QY 1491 GACCGAGTCAAGCGCGGCGGAGATGAGGCAACGAGCGGTCGCTGCTGCTGCTCATCTC 1550
 DB 1305 GCTGCAAGCTTCCGCGGAGCTGCGGCGGCAACGAGAGACATGTGATGCGCGGAGCG 1364
 QY 1551 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1609
 DB 1365 CCGCAAGCGGCTCTGAGGAGTGGCGGAGAAAGTCCCGCGGATTCGCGTGGAGTGGC 1424
 QY 1610 ACTGCTGTCAGGCGGCTGAGCTCGCGCGGAGTGAAGTTCGATTCAGAGGAGTTGG 1669
 DB 1425 TGGGCGCTGCGGAGGCTGAGCTTCGCGAGGCGCTGAAAGTTCGCGGAGCTGAGC 1484
 QY 1670 ACCCGTNNTCNCCGCTCCTCCTCAGGAGC 1699
 DB 1485 AGGCTGCGCGGCTGCTGCGGCAAGGTGC 1514

RESULT 14

US-09-252-991A-16126/c

; Sequence 16126, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-07-27

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16126

; LENGTH: 2295

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16126

Query Match 5.3%; Score 131.6; DB 4; Length 2295;
 Best Local Similarity 46.5%; Pred. No. 5.2e-16;
 Matches 572; Conservative 0; Mismatches 606; Indels 52; Gaps 5;

QY	471	CTTUNCGCTCGAGNCGGAGCTCGAAGAACTCGCTTCGCGTGGAGGTGTCGCGGCGGCAT	530
Db	1183	CTGTGCCATGCGCGCGGCGGTGGCGAGGCCCTGAGACAGCCATGATGCGGTGTAT	1124
QY	531	GACCATCGAGTCAACTCGCTCAGNCGGCGCACTCGAGNCTCGCTGTGTCTCGA	590
Db	1123	GCTGCTCAAGGTGAACAGGCTGAGCCGCGGGTCTTCGCGCATCGCGCGAAGGTATGCA	1064
QY	591	GGGCTCAACCACTTCCTCAACCAAGCATGACCCCATGTCGCCCTCGCGGCGACAT	650
Db	1063	CGGCTGTGATCGGCTGTATTAAGCCGAGGTTATCCGACATCCCGTGAAGGCTCGGT	1004
QY	651	CTGCGGCTCGGCGGACCTCTCCNCTCTCNTATCATGCGCGGCATCACTGTCACC	710
Db	1003	GGGTGCGTCCGCGGACCTGAGCCGCTGCGCCACATGCTCGGTGTATGAGGGAAG	944
QY	711	GGAANNCAAGGTACAGTNNNTACAGAGGAGANNAGAATATCATGNNCGCCGAGAGC	770
Db	943	CGGGGCGCGCATCGCGTG-----AGTGGCTGCCGCGCGGCAAGC	902
QY	771	GATCGGCTCTTNGTCTTCAGCCCGTGTCTCGCCCGAAGAGGTTCTGCTGTCTGT	830
Db	901	GCTGGGAGTGGCCGAGCTTGAGCCGCTACCTGCGCGCGAAGAGGCTCTGCTCTCT	842
QY	831	CAACGGCAAGGCGGTCTCGGCTCGATGAGCGACCTGTGCTCTGCAAGACACATGCT	890
Db	841	CAATGGCACCAAGTGTCCAGCGCTTAAGCGCTGCGGGTGTGTGAAGCCGAGACT	782
QY	891	CTGCTCTCTCGCAGGCGCTCACGGCTCTTACGAGTGAAGGCATGATGAGCCACGCGG	950
Db	781	GTTTCGCGCGCGCACCGCTGTGCGGAGGCTCAGGCTGAGGAGCATCTCGGTTTCGCGGC	722
QY	951	CTGTTTCAACCCNTTCTTCAAGAGTCAAGCGCCCTCAACCCGACCAATGAGAGTGC	1010
Db	721	GCGCTTCGATGCGCGATCAATGCGCGCGCGGCGACAGCTGGGCAATGAGCTGGCGC	662
QY	1011	GCGCAACATCGCAGCTCTCGAAGGAGCGNNGTTCGCCGTCACCAAGAGAGAGGT	1070
Db	661	GGCTTATCGGACTGTCTACCGCCAGCAGAGAGTGGCCGCTCCATGAGAAAGTGGCA	602
QY	1071	CAAGGTCAAGAGAGAGAGGATTTCTCGCGAGAGCGGTACCGGCTCGGACGCGGC	1130
Db	601	CAAGGT-----CCAGGACCTTATTCGCTGCGTTGCCAGCC	566
QY	1131	TGAGTGCCTCGCGCGCTGTGAGCAGATTTACGCCACGNGTCTCTGCTGTGA	1190
Db	565	GCAGGTGATGGGCGCTGTGCTGACCCCAATCGCCAGGCGCGAGGTCTCGAATGCA	506
QY	1191	GGCGGNCAGTGCAGACAGCAACCCGCTCATGAGTGCAGAAACAAMAGACCACCA	1250
Db	505	AGCT---AAGCGGTCTCGACACACCCTGTATTTGCGCGGAGGAGGATGATTC	449
QY	1251	CGCGGCAACTTCAGAGCGNCCGCTGTGCNAACAGATGAGAGAATCTGCTGTGNC	1310
Db	448	CGCGGCAACTTCAGCGCGCAACCGGTGGCATGCGCGCCAGAACCTGAGCTTGGCTT	389
QY	1311	CGCGCTGATCGGCAAGCTCAACTTACGCACTCACCGAATGCTCAAGCGCGATGAA	1370
Db	388	GGCGGAGATCGGTTCGCTGTGGAAGCGCGCATCTCGCTGATGATGACATGATTC	329
QY	1371	CGCGGACCTNCGTCTGTGCTGTGCTGCCGAGAACCTTGCCTCTCTATCATGCAAGG	1430
Db	328	GCAGTTCGCGCGTCTTCGTGTGGCCAGGG-----GGGTCAATTCGGGTTTATAT	275
QY	1431	CTTCGACATTCGCGGNGCGNTACACTTCGAGCTCGGACACTTCGCAACCCGATAC	1490
Db	274	CGCGCAGGTACCGCGCGCGCGCTGCGCAGAGGACAAACAAGCCCTTGACCATTCGGCGAG	215
QY	1491	GACCCACGTTCACCGCGGAGAGTGGGACACAGGCGGTCACTCGCTGCGNCTATCTC	1550
Db	214	CGTGCACACCTTCGCGACTTCGGCCCAACAAGAAACATATGTGTGATGGCGCGAAGC	155

Db 463 CAACGACAATCCGTTAATGATGTCTCCAGGGACATGGCTCTCCACGGCGGCAATTCCA 522
QY 1266 GGCGNCCGCTGTGGNAAACACGATGAGAAAGACTGCCTCGGCTCGC 1313
Db 523 GGGAACACCCATCGAGTTTCCATGACACATGCGAATCTCTTTGGC 570

Search completed: September 12, 2004, 04:26:57
Job time : 199.6 secs

Db	1	ATGAGCCCCCTCCMTGACTCGATCGGACACTCGTGGCCAAAGGACTCMNAACGGANTG	60
QY	61	CACGCCGNNCCGNCNNNGCNCNACAGGGCGCCACGTCCACNCTCNGNCGCGCCGNCNG	120
Db	61	CACGCCGNNCCGNCNNNGCNCNACAGGGCGCCACGTCCACNCTCNGNCGCGCCGNCNG	120
QY	121	GCTCGCTCTCCGCGACACCCAGNNAGACGACGTCCGACATCGNNAGNAGATCTCTCGCG	180
Db	121	GCTCGCTCTCTCCGCGACACCCAGNNAGACGACGTCCGACATCGNNAGNAGATCTCTCGCG	180
QY	181	ACCCACCCGNNACGNAAGNNNTGAACTCGAGCGGGTACACCTTCACCTTCGGNGACGTG	240
Db	181	ACCCACCCGNNACGNAAGNNNTGAACTCGAGCGGGTACACCTTCACCTTCGGNGACGTG	240
QY	241	TGGCGCCGCGNCGGCAAGGGCCGNCNGTCGCGCNCNCAGACANNCAGAGATCCGGCGCA	300
Db	241	TGGCGCCGCGNCGGCAAGGGCCGNCNGTCGCGCNCNCAGACANNCAGAGATCCGGCGCA	300
QY	301	AAGATGACAAANAGNGTCGAGTTCCTCCGNNCAGCTCNAACAACAGNTCTTACGGNGTC	360
Db	301	AAGATGACAAANAGNGTCGAGTTCCTCCGNNCAGCTCNAACAACAGNTCTTACGGNGTC	360
QY	361	AGGACTGGTTTGGGGGGCTCGGCCGACCCCGACTGAGATGCAATCTCGTCCAGAG	420
Db	361	AGGACTGGTTTGGGGGGCTCGGCCGACCCCGACTGAGATGCAATCTCGTCCAGAG	420
QY	421	GCNCTCCTGAGACACACACTCTGCGGTGNTCCCNACGTCGNTCGANTCTTTCGCGCTC	480
Db	421	GCNCTCCTGAGACACACACTCTGCGGTGNTCCCNACGTCGNTCGANTCTTTCGCGCTC	480
QY	481	GGNCGCGGCTTGAAGAATCGCTTCGCTGAGGTGTCGCGCGGCCATGACATCCGC	540
Db	481	GGNCGCGGCTTGAAGAATCGCTTCGCTGAGGTGTCGCGCGGCCATGACATCCGC	540
QY	541	GTCAATCGCTCANCNGCGGCACTGGGNCGTGCGCTCGTCTCGAGGCGGTCAAC	600
Db	541	GTCAATCGCTCANCNGCGGCACTGGGNCGTGCGCTCGTCTCGAGGCGGTCAAC	600
QY	601	AACCTTCCTAACACAGGAGTACAGCCCATGGTCCCTCCGCGGACCATCTCGGCGTGC	660
Db	601	AACCTTCCTAACACAGGAGTACAGCCCATGGTCCCTCCGCGGACCATCTCGGCGTGC	660
QY	661	GGGAGACTCCGNCCTCTMTAATAGCGCGGCCATACACGCTATACCCGGACNNAG	720
Db	661	GGGAGACTCCGNCCTCTMTAATAGCGCGGCCATACACGCTATACCCGGACNNAG	720
QY	721	GTNCAAGTNTTACAGAGGCAANNAGAAAGTATGNTCCCGCGAGGCGATTCGCGTTC	780
Db	721	GTNCAAGTNTTACAGAGGCAANNAGAAAGTATGNTCCCGCGAGGCGATTCGCGTTC	780
QY	781	TTNAGTCTCGAGCCCTGTCCTTGCGCCCGAAGAGGGTTCGCTTCGTCAACGGCACG	840
Db	781	TTNAGTCTCGAGCCCTGTCCTTGCGCCCGAAGAGGGTTCGCTTCGTCAACGGCACG	840
QY	841	GGCGTTCGCGTCGATAGGCAACCTGCGCTGACAGACACAATGCTTCGCTCCTC	900
Db	841	GGCGTTCGCGTCGATAGGCAACCTGCGCTGACAGACACAATGCTTCGCTCCTC	900
QY	901	TGCGAGGCGCTCAGGACTCTNACGGTGAAGGCATGGTCGCGCACCGCGACTCTTCCAC	960
Db	901	TGCGAGGCGCTCAGGACTCTNACGGTGAAGGCATGGTCGCGCACCGCGACTCTTCCAC	960
QY	961	CCNTTCCTCAGAGAGTACGGCGGCTCACCACCCAGATGAGAGTTCGCGGCAATC	1020
Db	961	CCNTTCCTCAGAGAGTACGGCGGCTCACCACCCAGATGAGAGTTCGCGGCAATC	1020
QY	1021	CGCAGGCTCTGAGGCGAGCANNGTTCGCTCCACACAGAGGAGGATCAAGSTCAAG	1080
Db	1021	CGCAGGCTCTGAGGCGAGCANNGTTCGCTCCACACAGAGGAGGATCAAGSTCAAG	1080
QY	1081	GACGAGAGGAGATTTCTCCGCCAAGACCGCTACCCGTCGCAACGTCTAGTGGTCT	1140
Db	1081	GACGAGAGGAGATTTCTCCGCCAAGACCGCTACCCGTCGCAACGTCTAGTGGTCT	1140

QY	1144	GGCCCGCTGTCAGAGCAKATGATTTACACGCGCACGCGNCTCTGCGTGAAGCCGCGNAG	1200
Db	1141	GGCCCGCTGTCAGAGCAATGATTTACGCGCCACGCGNCTCTGCGTGAAGCCGCGNAG	1200
QY	1201	TCGACGACCGACCAACCCGCTCATCGAGCTGAGAAACAAGNAGCCACACCGCGGCAAC	1260
Db	1201	TCGACGACCGACCAACCCGCTCATCGAGCTGAGAAACAAGNAGCCACACCGCGGCAAC	1260
QY	1261	TTCCAGGCGAGCGCTGTGCGNACAAGATGAGAGAAAGACTCGGCTGCGNCTGCGCTGTATC	1320
Db	1261	TTCCAGGCGAGCGCTGTGCGNACAAGATGAGAGAAAGACTCGGCTGCGNCTGCGCTGTATC	1320
QY	1321	GGCAAGCTCAACTTCACGAGCTCACCGAGATGCTCAAAGCGCGGATGACACCGCGGCTTN	1380
Db	1321	GGCAAGCTCAACTTCACGAGCTCACCGAGATGCTCAAAGCGCGGATGACACCGCGGCTTN	1380
QY	1381	CNCTCCGCGCTGCGCTGCGAGGACCCNTGCTCTCTCTATCACTGCAAGGCGCTGACATT	1440
Db	1381	CNCTCTCGCTGCGCTGCGAGGACCCNTGCTCTCTCTATCACTGCAAGGCGCTGACATT	1440
QY	1441	GCCGCGNCGNCTNACACTTCGAGCTCGANCACTTGGCAACCGCGGTNACGACCGATC	1500
Db	1441	GCCGCGNCGNCTNACACTTCGAGCTCGANCACTTGGCAACCGCGGTNACGACCGATC	1500
QY	1501	CAGCGCGNAGATGAGGACAACAGGCGCGTCACTGCTGAGNCTATCTCGGCGCGCGC	1560
Db	1501	CAGCGCGNAGATGAGGACAACAGGCGCGTCACTGCTGAGNCTATCTCGGCGCGCGC	1560
QY	1561	ACNCGCGAGGCGCAAGACGTCTTTTCTCTCTCTCTCGGCAACCACTCTACTGGGTNCTC	1620
Db	1561	ACNCGCGAGGCGCAAGACGTCTTTTCTCTCTCTCTCGGCAACCACTCTACTGGGTNCTC	1620
QY	1621	CAGCGCGTGAACCTCGCGCGATGAGATTGAGATTCAAGAAAGCATTTGACCCGATNNTC	1680
Db	1621	CAGCGCGTGAACCTCGCGCGATGAGATTGAGATTCAAGAAAGCATTTGACCCGATNNTC	1680
QY	1681	NCNNGCTCNTNAGAGACACTTTTGGCNCNGCCCTCTGACGCGNNNNNNNNNNNACAA	1740
Db	1681	NCNNGCTCNTNAGAGACACTTTTGGCNCNGCCCTCTGACGCGNNNNNNNNNNNACAA	1740
QY	1741	CTCGNGGACAAAGSTCAACAAGNCGCTCNACAGCNCNCTGAGCAGACCACTCGTACGAC	1800
Db	1741	CTCGNGGACAAAGSTCAACAAGNCGCTCNACAGCNCNCTGAGCAGACCACTCGTACGAC	1800
QY	1801	CTCGAGCGCGCTGCGACGACGCGCTTCTGTTCTGCGACCGGACCGCTGTCGAGNNNNN	1860
Db	1801	CTCGAGCGCGCTGCGACGACGCGCTTCTGTTCTGCGACCGGACCGCTGTCGAGNNNNN	1860
QY	1861	NNGTCCTCGCNCNNNNCCANNAGSTGTGCTGCGNCGGTCAACGCTGGAAGGTGCGCT	1920
Db	1861	NNGTCCTCGCNCNNNNCCANNAGSTGTGCTGCGNCGGTCAACGCTGGAAGGTGCGCT	1920
QY	1921	CCGCGCGAAGAGGCATCTGCTCAAGCGCGNAGTCCGACNCCTTCTGAGNCGCGNCCGT	1980
Db	1921	CCGCGCGAAGAGGCATCTGCTCAAGCGCGNAGTCCGACNCCTTCTGAGNCGCGNCCGT	1980
QY	1981	CGTGTGTGTGAGCCGCGGCTCNGTACCTCTGCGCGGACGCGGCTCTGTACTGTGTCG	2040
Db	1981	CGTGTGTGTGAGCCGCGGCTCNGTACCTCTGCGCGGACGCGGCTCTGTACTGTGTCG	2040
QY	2041	TCGCGCGAAGAGCTCGGCGCTCAAGGCGCGCGCGCGGAGAGTCTTCTCGGCAAGAGAGG	2100
Db	2041	TCGCGCGAAGAGCTCGGCGCTCAAGGCGCGCGCGCGGAGAGTCTTCTCGGCAAGAGAGG	2100
QY	2101	TGACGATCGGCACCAACGCTCTCCGCGATCTACGAGGCATCAAGNCGGCGCATCAAC	2160
Db	2101	TGACGATCGGCACCAACGCTCTCCGCGATCTACGAGGCATCAAGNCGGCGCATCAAC	2160
QY	2161	ACGCTCTCTGCAAGAGCTCGCNTAGNNCNCNNNCNANNCTCGCNTNNNNCCANNNC	2220
Db	2161	ACGCTCTCTGCAAGAGCTCGCNTAGNNNCNNNCNANNCTCGCNTNNNNCCANNNC	2220

QY 2221 NNNCCNNNNNNCTTTNNNNNTGGNTTCTNTGNCNNNNNGAANTTNNCNCNNNNNN 2280
DB 2221 NNNCCNNNNNNNNCTTTNNNNNTGGNTTCTNTGNCNNNNNGAANTTNNCNCNNNNNN 2280
QY 2281 TNNNNCTNNCTNNCTGNCNNNNNANNGTCTNTNNNNCTNNNNNTNNNNNNNNNN 2340
DB 2281 TNNNNCTNNCTNNCTGNCNNNNNANNGTCTNTNNNNCTNNNNNTNNNNNNNNNN 2340
QY 2341 NGTNNCCANNNAACNTNNNNNNNNNNNNNNNGGANNNGAANTTNNNNNNNNNN 2400
DB 2341 NGTNNCCANNNAACNTNNNNNNNNNNNNNNNGGANNNGAANTTNNNNNNNNNN 2400
QY 2401 NNNNAAAA 2408
DB 2401 NNNNAAAA 2408
RESULT 2
US-09-939-408A-29
Sequence 29, Application US/09939408A
Patent No. US20020102712A1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
APPLICANT: Koester, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSA
CURRENT APPLICATION NUMBER: US/09/939,408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 2163
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Preferred
NAME/KEY: CDS
LOCATION: (1)..(2163)
OTHER INFORMATION:
US-09-939-408A-29
Query Match 78.5%; Score 1944; DB 9; Length 2163;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2037; Conservative 0; Mismatches 126; Indels 23; Gaps 7;
QY 1 ATGGCCCTCCCTGCTGATCGATGGCGACCTCGAGTGGCGCAAGGCTCNCNAAAGGATG 60
DB 1 ATGGCCCTCCCTGCTGATCGATGGCGACCTCGAGTGGCGCAAGGCTCNCNAAAGGATG 60
QY 61 CACGCCGNNCCGNCNNNNNGCNCNACGAGCGCCACGTCACNCTCNGNCGCCGNCNG 120
DB 61 CACGCCGCTCCGHCACAGTCGSCMAAGGCGCCACGTCACNCTCNGNCGCCGNCNG 120
QY 121 GCTCGCTCTCCCGACACACCGGANNNGAGCGAGCTGAGATGTTGAGAAAGTCTTCCGCG 180
DB 121 GCTCGCTCTCCCGCG-CCACCCAGVHAGCGAGCTGAGATGTTGAGAAAGTCTTCCGCG 178
QY 181 ACCCCACCGNNAAGNAGNNNTGAACTGACGAGGTACACCTCGAGAGCGTGC 240
DB 179 ACCCCACCGAC---GACGMSVTCGAATCGACGAGGTACACCTCGAGAGCGTGC 235
QY 241 TCGGCGCGCGCGCAAGGCGCGCNCNCTCCGCTGNCNCAAGNCGAGAGATCCGCGCA 300
DB 236 TCGGCGCGCGCGCAAGGCGCGCHCBGTCGCGTTC-CAGACAGNCGAGAGATCCGCGCA 294
QY 301 AAGATGCAANANAGTGAAGTTCTCCGNNCNCAGCTCNAACAAGAGTGTACGAGTGC 360

DB 295 AAGATGCAAAVABGTTGAGATTCCTCCGDBCBAGCTCBACAAACABGTTACGGHGTG 354
QY 361 ACGACTGGTTTCGCGGCTCGGCGGACACCCGACGACGAGTGAATGCAATCGCTCCAGAAG 420
DB 355 ACGACTGGTTTCGCGGCTCGGCGGACACCCGACGACGAGTGAATGCAATCGCTCCAGAAG 414
QY 421 GCGCTCTCGAGCACACGACTGCGGTGTNCTCCCAACGTCGANTGCAATCTCTNCGCTC 480
DB 415 GCBCTCTCGAGCACACGACTGCGGTGTNCTCCCAACGTCGANTGCAATCTCTNCGCTC 474
QY 481 GGNCGCGGCTCGAAGACTCGCTCCGCTCGAGGTGCTCGCGGCGGACATCGACATCGCG 540
DB 475 GGHCGCGGCTCGAAGACTCGCTCCGCTCGAGGTGCTCGCGGCGGACATCGACATCGCG 534
QY 541 GTCAACTGCTCAACGCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGG 600
DB 535 GTCAACTGCTCAACGCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGG 594
QY 601 AACTCTCTCAACGCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGG 660
DB 595 AACTCTCTCAACGCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGG 654
QY 661 GCGGACTCTCCGCT 720
DB 655 GCGGACTCTCCGCT 714
QY 721 GTNCAAGTNNNNNCAAGGAGGANNNGAAGATCATGTTNCGCGCGGAGGAGTCCGCTC 780
DB 715 GTNCAAGTNNNNNCAAGGAGGANNNGAAGATCATGTTNCGCGCGGAGGAGTCCGCTC 774
QY 781 TTNAGTCTCGAGCCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
DB 775 TTBGCTCTCGAGCCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 834
QY 841 GCGCTCTCGGCTCGAGTGGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
DB 835 GCGCTCTCGGCTCGAGTGGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 894
QY 901 TCGAGGCGCTCAAGGCTCTNACGAGTGGGACCTCTCTCTCTCTCTCTCTCTCTCTCT 960
DB 895 TCGAGGCGCTCAAGGCTCTNACGAGTGGGACCTCTCTCTCTCTCTCTCTCTCTCTCT 954
QY 961 CCNTTCTCTCAAGAGTCAAGGCTCTCAAGGCTCTCAAGGCTCTCAAGGCTCTCAAG 1020
DB 955 CCNTTCTCTCAAGAGTCAAGGCTCTCAAGGCTCTCAAGGCTCTCAAGGCTCTCAAG 1014
QY 1021 CGGAGCTCTCTCGAGGAGGAGGNNGTTTCCGCTCCACGAGGAGGAGGTTCAAGTCAAG 1080
DB 1015 CGGAGCTCTCTCGAGGAGGAGGNNGTTTCCGCTCCACGAGGAGGAGGTTCAAGTCAAG 1074
QY 1081 GACGACGAGGAGGATCTCCGCGACGAGTACCGGCTCCGACGAGTCCGCTCAAGTGCCTC 1140
DB 1075 GACGACGAGGAGGATCTCCGCGACGAGTACCGGCTCCGACGAGTCCGCTCAAGTGCCTC 1134
QY 1141 GCGCGCTCTCTGAGGAGTATTTACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
DB 1135 GCGCGCTCTCTGAGGAGTATTTACGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1191
QY 1201 TCGAGACCGAACAACCGCTCATGAGTGGAGAAAGAAAGAACCCACGAGGCGCAAC 1260
DB 1192 TCGAGACCGAACAACCGCTCATGAGTGGAGAAAGAAAGAACCCACGAGGCGCAAC 1251
QY 1261 TTCAAGGCGGCGGCTGTGCAACAAGATGAGAAAGTCCGCTCGGCTCTGCGCTGATC 1320
DB 1252 TTCAAGGCGGCGGCTGTGCAACAAGATGAGAAAGTCCGCTCGGCTCTGCGCTGATC 1311
QY 1321 GCGAAGCTCACTTCAAGCAAGTCAAGCAAGTCAAGCAAGTCAAGCAAGTCAAGCAAG 1380
DB 1312 GCGAAGCTCACTTCAAGCAAGTCAAGCAAGTCAAGCAAGTCAAGCAAGTCAAGCAAG 1371
QY 1381 CCGTCTGCTCTGCTCGAGAGACCCNTGCTCTCTATCACTGCAAGGCTCGACAT 1440

Db	137	CGBTCTGGCTCGCTGCGAGGAGCCGBCGCTCTCATATGACGAAAGGCGCTCCACANTT	1431
Qy	1441	GCCGCGNCGNCTACACTTTCGAGGTCTGGNCACTTTCGCAACCCGGTNAGACCAAGTC	1500
Db	1432	GCCGCGGCGGCBTACACTTCGAGGTCTGGNCACTTTCGCAACCCGGTBAGACCAAGTC	1491
Qy	1501	CAGCCGCGNAGATGGGACCAAGGCGGTCAACTGCTGCGNCTCATCTCGGNGCGCG	1560
Db	1492	CAGCCGCGGCBGATGGGCAACCAAGGCGGTCAACTGCTGCGBCTCATCTCGGCBGCGCG	1551
Qy	1561	ACNCGCCGAGGCCAAGCAAGTCCTTTCTCTCTCTCTGCGCAACCACTTAATGCGTCTC	1620
Db	1552	ACBGCAGGAGGCCAAGCAAGTCCTTTCTCTCTCTCTGCGCAACCACTTAATGCGTCTC	1611
Qy	1621	CAGGCGGTGACCTCGCGGCAATGAGAGTTCAGATTCAAGAGAGTTCGACCCGATNTTC	1680
Db	1612	CAGGCGGTGACCTCGCGGCAATGAGAGTTCAGATTCAAGAGAGTTCGACCCGATNTTC	1671
Qy	1681	NCANCGCTCNTNAGACGACTTTTGGCNCNGCCCTCGACGCGNNNNNNNNNNNACAA	1740
Db	1672	VCBDCGCTCHTCAAGAGACTTTTGGCNCNGCCCTCGACGCG-----NACGAA	1739
Qy	1741	CTGNGNCAAAAGTCAACAAAGTCTTAAACAAGGNTCGAGACACAACTTCGATGAC	1800
Db	1720	CTGNGNCAAAAGTCAACAAAGTCTTAAACAAGGNTCGAGACACAACTTCGATGAC	1779
Qy	1801	CTGAGACCGCGCTGCGACAGAGCGCTTCTGTTTCGCGACCGGCAACGTCGCGANNNNN	1860
Db	1780	CTGAGACCGCGCTGCGACAGAGCGCTTCTGTTTCGCGACCGGCAACGTCGCGCTC	1839
Qy	1861	NNGTCTCGCCNNNNNCCANNAGGTCTGCTGCGNCGCGTCAACGCTTGGAAAGTCGCT	1920
Db	1840	TCGTCTCTCCGCTGCTGCA--AGGTCTGCTGCGGCGGCTCAACGCTTGGAAAGTCGCT	1897
Qy	1921	CGCGGAGAAAGGCATCTGCGTCAACGCGCNANAGTCGCGCAACNCTTCTGAGNGGNCCT	1980
Db	1898	CGCGGAGAAAGGCATCTGCGTCAACGCGCAVGTCCGCAHCCTTCTGAGGCGGBCGT	1957
Qy	1981	CGTGTGCTGCGCGCGCGCTCNCGTAACTCTTGCGCGGCAAGCGCGTCTGTAAGTTCG	2040
Db	1958	CGTGTGCTGCGCGCGCGCTCDDGTAACTCTTGCGCGGCAAGCGCGTCTGTAAGTTCG	2017
Qy	2041	TCCGCGAGAGGCTCGGCGTCAAGAGCGCGCGCGGCGACGTTCTTCGCGCAGCAGAGG	2100
Db	2018	TCCGCGAGAGGCTCGGCGTCAAGAGCGCGCGCGGCGACGTTCTTCGCGCAGCAGAGG	2077
Qy	2101	TGACGATCGGACCAAGGCTCCGCACTTACGAGGCGATCAAGNNCGGCGNATCAACC	2160
Db	2078	TGACGATCGGACCAAGGCTTCCGCACTTACGAGGCGATCAAGGVCGGRHCAATCAACC	2137
Qy	2161	ACGTCTCTGTCAAAGTCTCGCNTAG	2186
Db	2138	ACGTCTCTGTCAAAGTCTCGCNTAG	2163
RESULT 3			
US-09-939-408A-18			
Sequence 18, Application US/09939408A			
Patent No. US20020102712A1			
GENERAL INFORMATION:			
APPLICANT: Yoshida, Roberta			
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and			
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and			
FILE OF INVENTION: Using Same			
FILE REFERENCE: 29479/500NSCA			
CURRENT APPLICATION NUMBER: US/09/939, 408A			
CURRENT FILING DATE: 2001-08-24			
PRIOR APPLICATION NUMBER: US 09/624, 693			
PRIOR FILING DATE: 2000-07-24			
PRIOR APPLICATION NUMBER: PCT/US01/23270			
PRIOR FILING DATE: 2001-07-24			
NUMBER OF SEQ ID NOS: 30			

```

; SOFTWARE: PatentIn Ver. 2.0.0
; SEQ ID NO 18
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Rhodotorula torul
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(2148)
US-09-939-408A-18

```

Query Match 68.7%; Score 1699.4; -DB 9; Length 2439;
Best Local Similarity 79.4%; Pred. No. 0;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

[illegible]

[illegible]

```

Ox      1  ATGGCCCCCTCCNTGACTGATCGCAGACCTGCAGTCCAGTCCGCAAGCGCTCNCNCAACGAGTNG 60
Oy      1981 CGTCGTCGTCGCGCCCGCTCNCGACTCTCGCGCGCAGCGCGCTCGTACTGGTTCG 2040
Db      1946 CGACCTCGTCGCGCGCGCTCTCGTACTCTCTCGCGCGACTCAGATCCCTTACGCTTCG 2005
Oy      2041 TCCGCGAGGAGCTCGGCGCTCAAGCGCGCGCGCGCGACTCTTCTCTCGGCAACGAGAG 2100
Db      2006 TCCGCGAGGAGGCTTGGCGCTCAAGGCCCGCGCGGAGACGCTTCTCTCGCAAGAGAGG 2065
Oy      2101 TGACGATCGGCACCAAGCTCTCCCGCATCTACGAGGCGATCAAGNNCGGCNCGATCAAC 2160
Db      2066 TGACGATCGGCTCGAAGCTCTTCCAAGATCTACGAGGCCATCAAGTGCGGCAAGATCAACA 2125
Oy      2161 AGCTCTCTGTCGAAGTGTCTGCGCTAGNNNCNANNNCNANNCTGCGCTNNNNNCNANNNC 2220
Db      2126 AGCTCTCTCTCAAGATGCTCGCTTAGACACTTCCCACTCTGCACTCCCTTCATPACC 2185
Oy      2221 NNNNCNNNNNNNCNTTNGNNNTGNNNTGNNNTGNNNTGNNNNCGGANTTNNCNNNNNNNN 2280
Db      2186 TATCCCGCGCTGCACTTAGAGCTGCTTCTTGTGCGACTCGGATCTCGCATCTGCTTCT 2245
Oy      2281 TNNNNCNTTNNCTNNCTCNANNANNCNNCTNNNNNNCTNNNGNTNNNNNNNNNN 2340
Db      2246 TGCTTCTTGCGCTGCTCTC -TAGACGCTGTCCGATTACTGGAAGTTGTGAATPAAAG 2304
Oy      2341 NCTNNNCAANNACNCTNNNNNNNANNCGNANNGANTTNNAGNTTNGNNGNNNNNNN 2400
Db      2305 AGTACCCATCCACGATCCGATTAATCAAGGAGAAATCTAGCGTTGCGGAGCTTCTTG 2364
Oy      2401 NNNNAAAA 2408
Db      2365 CGCATPAA 2372

RESULT 4
US-09-939-408A-12
/ Sequence 12, Application US/09939408A
/ Patent No. US20020102712A1
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ APPLICANT: Koelstra, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ TITLE OF INVENTION: Using Same
/ FILE REFERENCE: 29479/500NSCA
/ CURRENT APPLICATION NUMBER: US/09/939,408A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 09/624,693
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ PRIOR FILING DATE: 2001-07-24
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 2419
/ TYPE: DNA
/ ORGANISM: Rhodotorula graminis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (37)..(2196)
/ NAME/KEY: modified_base
/ LOCATION: (494)
/ OTHER INFORMATION: Other information: Y = t or c
/ NAME/KEY: unsure
/ LOCATION: (493)..(495)
/ OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408A-12

Query Match          65.8%; Score 1629; DB 9; Length 2419;
Best Local Similarity 78.0%; Pred. No. 0;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

```


QY 2221 NNNCCNNNNNNCTTTNNNNNTGNNNTGNNNNNNNNNNCGGANNNTNNNNNNNN 2280
Db 2233 ACCCCAAGACCAAGCTTTTCAGCGCTGTGTGTCAGAAACGAGCTTCTTCATACACA 2292
QY 2281 TNNNNCTNNCTNNCTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
Db 2293 TGTGCTTACTCTCTGCGCGGTGCATCGTCTCTAGTTCTTTCGTATCCGCGTCTCTC 2352
QY 2341 NGTNN 2400
Db 2353 GGTGCTGA-GTACACGTATATAGAGCTGGAATGATGCAAGTCTTGAATCAAAAA 2411
QY 2401 NNNNNAAA 2408
Db 2412 AAAAAAAA 2419

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

Query Match 65.7%; Score 1626.4; DB 9; Length 2151;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGAGCCCTCCCTCGACTCGATCGGACCTCGANTGCGCAACGAGNTGNNNAGGANTG 60
Db 1 ATGGCAACCTCGCTCGACTCGATCGGACCTCGANTGCGCAACGAGNTGNNNAGGANTG 60
QY 61 CAGCGCGNNCCGNCNNNNNGCNCNACGAGCGCCGCAAGTCCACNCTGNCNGCGCGNCGNG 120
Db 61 CAGGCTGT-----CAATGGCGCTCGACCACTC-----GAGTGGCAG 100
QY 121 GCTGCTCTCCCGGACACACCCAGNNNAGCGAGCTGACATCTGTTNAGNAGATCTTCGCGCG 180
Db 101 GCTGCACTCTGCGCAACCCAGGTGACGAGGTGACATCTGTCAAGAAATGCTTCGCGCG 160
QY 181 ACCCAACCGNNAAGNAGNNNTGAACTGAGCGGTACACCTTCACTCCGAGNAGCTCG 240
Db 161 CGCGAACCGACTGA---CGCTGAACCTGACGAGGTGACCTGCTCAACTCGGAGAGCTCG 217
QY 241 TCGGCGCGCGCGCAAGGGCGCGNCNGTCCGCTCNCAGACAGNAGAGAGATCCGCGCA 300
Db 218 TCTGCGCGCGGAGAAAGGGGAGGCGCTGTCCGCTCAAGAGACAG-GAGAGAGATCCGCTCA 276
QY 301 AAGATGACAAANAGNGTGAAGTCTCCGNNNCAGCTGNAACAAGNGTCAAGGANTG 360
Db 277 AAGATGACAAATCGGTGAGTCTTTCGCGCTGCAACTCTCATAGAGGTCTTCAAGGCTC 336
QY 361 AGCAGTGTGTTGCGCGGCTCGCGCGACACCGAGCTGAGAGATGCAATCTCGCTCCAGAG 420
Db 337 AGCAGTGTGATTTGGCGGATCCGAGACACCGGACCGAGAGAGCGCATCTCGCTCCAGAG 396

QY 421 GCNCTCTTGAGACCAAGCTTTCGCGGTGNTCTCCNAGGTGANTGANTCTTCCGCTC 480
Db 397 GCTCTCTGAGAGACCAAGCTTTCGCGGTGNTCTCCNAGGTGANTGANTCTTCCGCTC 456
QY 481 GAGCGGAGCTCGAGAACTCGCTTCGCTCGAGGTGTCGCGGCGGACCATGCAATCCG 540
Db 457 GAGCGGAGCTCGAGAACTCGCTTCGCTCGAGGTGTCGCGGCGGACCATGCAATCCG 516
QY 541 GTCACTGCTCAACNCGCGGCACTTTCGCGGTGTCGCTGTCGCTTTCGAGGCTCAC 600
Db 517 GTCAAGCTTGAACCGCGGCACTTTCGCGGTGTCGCTGTCGCTTTCGAGGCTCAC 576
QY 601 AACTCTCTCAACCGCGGCACTTTCGCGGTGTCGCTGTCGCTTTCGAGGCTCAC 660
Db 577 AACTCTCTCAACCGCGGCACTTTCGCGGTGTCGCTGTCGCTTTCGAGGCTCAC 636
QY 661 GAGCACTTCTCCGCTCTCTCAATGCTGCGCGGCACTTTCGCGGTGTCGCTTTCGAGGCTCAC 720
Db 637 GAGCACTTCTCTCTCTCTCTCAATGCTGCGCGGCACTTTCGCGGTGTCGCTTTCGAGGCTCAC 696
QY 721 GTNCACTGNTNCAAGAGGCGGCAAGAGATCATGTCGCGCGGAGGCGATTCGCTC 780
Db 697 GTNCACTGNTNCAAGAGGCGGCAAGAGATCATGTCGCGCGGAGGCGATTCGCTC 756
QY 781 TTNGTCTGAGCGCGCTGCTCTCGCGCGGCAAGAGGCTGCTGCTGCTCAACGCGCAG 840
Db 757 TTNAACTGAGCGCGCTGCTCTCGCGCGGCAAGAGGCTGCTGCTGCTCAACGCGCAG 816
QY 841 GCGGTTCGCGCTGATGAGGCACTTTCGCTGCAAGAGGCAACATCTCTGCTCTC 900
Db 817 GCGGTTCGAGCATGATGAGGCACTTTCGCTGCAAGAGGCAACATCTCTGCTCTC 876
QY 901 TGGCAGGCGCTCAAGCGCTCTNAGCGGCGGCAAGGCGGCAAGGCGGCGGCTTCCAC 960
Db 877 TGGCAGGCGCTCAAGCGCTCTNAGCGGCGGCAAGGCGGCAAGGCGGCGGCTTCCAC 936
QY 961 CCNTTCTCAAGCGCTCAAGCGGCGGCAAGGCGGCAAGGCGGCGGCTTCCAC 1020
Db 937 CCNTTCTCAAGCGCTCAAGCGGCGGCAAGGCGGCAAGGCGGCGGCTTCCAC 996
QY 1021 CGGAGCTCTGAGAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1080
Db 997 CGGAGCTCTGAGAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1056
QY 1081 GAGGAGGAGGATCTTCGCGGAGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1140
Db 1057 GAGGAGGAGGATCTTCGCGGAGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1116
QY 1141 GAGCGCTCTGAGAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1200
Db 1117 GAGCGCTCTGAGAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1176
QY 1201 TGGAGAGCGGCAACCGGCTCTATGAGCGGCGGCAAGGCGGCGGCTTCCAC 1260
Db 1177 TGGAGAGCGGCAACCGGCTCTATGAGCGGCGGCAAGGCGGCGGCTTCCAC 1236
QY 1261 TTCAGGCGGCGGCTCTATGAGCGGCGGCAAGGCGGCGGCTTCCAC 1320
Db 1237 TTCAGGCGGCGGCTCTATGAGCGGCGGCAAGGCGGCGGCTTCCAC 1296
QY 1321 GAGGAGCTCACTTCAAGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCCAC 1380
Db 1297 GAGGAGCTCACTTCAAGAGGCTTCAAGGAGGCTTCAAGGAGGCTTCCAC 1356
QY 1381 CCNTCTGCTGCTGAGGAGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1440
Db 1357 CCNTCTGCTGCTGAGGAGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1416
QY 1441 GCGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1500
Db 1417 GCGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCAAGGCGGCGGCTTCCAC 1476

Db 101 GTGACACTGCTCCCAACCCAGGTCAAGAGGTGATGTAAGAAATGCTGCGG 160
 Qy 181 ACCCCACCGNNAAGNACGANNNTGAACTCGAGGGTACACCTTCACCTCGGAGAGCTCG 240
 Db 161 GCGCCACGACCTCGA---CGCTCGAACTCGAAGGTACTGCTCAACTCGAGAGCTCG 217
 Qy 241 TCGAGCGCGCGCGCAAGGGCGCGCNGTCCGCTGTCNCAAGACNCAAGATCCGCGCA 300
 Db 218 TCTCGCGCGCGGGAAGGGCAAGGCTGTCCGCTCAAGACAG-CGAGGATTCGCTCA 276
 Qy 301 AAGATCGACAAANAGNGTGAAGTCTCCGANNONCAGCTCNACAAAGNGTCTAGAGGTG 360
 Db 277 AAGATTGACAAATCGGTGAGTTCTTGCGCTCGAACTCTCCATGAGCGTCTACGGGCTC 336
 Qy 361 ACGACTGTTTGGCGGCTCGGCGCAACCCGAGTGAAGATGCAATCTGCTCGAAG 420
 Db 337 ACGACTGATTTGGCGGATTCGCAACAACCGCAACCGAGGACGCACTCGCTCGAAG 396
 Qy 421 GGNCTCTCTGAGACCAAGCTCTGCGGCTGTCNCCNAGTCGATGATCTTNGCTC 480
 Db 397 GCTCTCTCGACACAGCTCTGCGGTGTCTCCCTTGTGTTGATCGTCTCGCTC 456
 Qy 481 GGNCGCGGCTCGAGAACTCGCTTCGCTCGAGGTCTCGCGGCGCATGACATCGC 540
 Db 457 GCGCGGCTCTGAGAACTCGCTTCCTTCGAGGTGTTGCGCGCGCATGACAAATCGC 516
 Qy 541 GTCAACTCTCTCAACGCGGCGCACTCGCGNGTTCGCTCTGCTCGAGGCGCTCAC 600
 Db 517 GTCAACAGCTTACCGCGCGCACTCGCTGTCCGCTGTGCTCTCGAGGCGCTCAC 576
 Qy 601 AACTTCCTCAACCAAGGAGTCAACCCCAATCGTCCCGCGCGGCAACATTCGCGCTC 660
 Db 577 AACTTCCTCAACCAAGGAGTCAACCCCAATCGTCCCGCGCGGCAACATTCGCGCTC 636
 Qy 661 GCGGACTCTCTCCGCTCTCTCTCAATCGCGCGCGCATCACCGTCAACCGGACNNAG 720
 Db 637 GCGGACTCTCTCTCTCTCTCTCAATGAGCGCGCATCAAGGATCAACCGGACGAG 696
 Qy 721 GTNCAAGTNNNCAAGAGGCGCANNGAAGATCATGTCGCGCGGAGGAGATCGGCTC 780
 Db 697 GTGCAAGTGTGTCACAGGAGGCAAGAGATCTGTAGCGCGGAGGAGCATGAGCTC 756
 Qy 781 TTNAGTCTGAGCGCGCTGCTCTCGCGCGCGGAGGAGGCTCGGTCTGTCAACGAG 840
 Db 757 TTTAACTCTGAGCGCGCTCTCTCTCGCGCGGAGGAGGCTCGGTCTGTCAACGAG 816
 Qy 841 GCGGCTCTCGGCTCGATGCGGAGCCTCTGCTGTCAAGCGACACATGCTTCTGCTCTC 900
 Db 817 GCGGCTCTGAGCATGATGCGGAGCCTCTGCTGTCAAGCGACACATGCTTCTGCTCTC 876
 Qy 901 TCGCAAGCGCTCAAGGCTCTMAAGGTTCAGAGGCGATGTCGCAAGCGGCGCTTCGAC 960
 Db 877 TCGCAAGTGTGTCAGGCGCATGAGGCTGAGAGGATGTCGCAAGCGGCGCTTCGAC 936
 Qy 961 CANTTCTCTCAAGCGTCAAGCGGCTCTCAAGCGGCGCATGAGGCTTCGCGCAACATC 1020
 Db 937 CCTTCTCTCAAGCGTCAAGCGGCTCTCAAGCGGCGCATGAGGCTTCGCGCAACATC 996
 Qy 1021 CGCAAGCTCTCTCAAGGCGCANNGTTCGCTGTCAAGCGAGAGGAGATCAAGTCAAG 1080
 Db 997 CGCAAGCTCTCTCAAGGAGAACCGCTTGTGTCAAGCGAGAGGAGATCAAGTCAAG 1056
 Qy 1081 GAGGAGAGGAGGATTCGTCGCGCGAGGAGCGGCTACCGGCTCGGAGTTCGAGTGTG 1140
 Db 1057 GAGGAGAGGAGGATTCGTCGCGAGGAGCGGCTACCGGCTTCGAGTTCGAGTGTG 1116
 Qy 1141 GCGCGGCTCTCTCAAGCATGATTCAGCGGCAACGCGCTCTCTCTCGAGGCGGAG 1200
 Db 1117 GCGCGGCTCTCTCAAGCATGATTCAGCGGCAACGCGCTCTCTCTCAAGGCGGAG 1176
 Qy 1201 TCGAGACCGGACACCGGCTCTCTCAAGCATGATTCAGGAGGAGGAGGAGGAGGAG 1260

Db 1177 TCGAGACCGGACCAACCTCTCTATCGAGGTGAGAGAAAGATTCGACAGAGGCGGCAAT 1236
 Qy 1261 TTTCAAGGACCGGCTGTGCGNAAACAGATGAGAAAGATCTGCTCGNCTCGGCTGATC 1320
 Db 1237 TTTCAAGGCTGCGGCTGTGCGCAACCATGAGAGATCTGCTCGGCTCGGCTGATC 1296
 Qy 1321 GCGAAGCTCAATTCAGAGGCTCACGAGATGCTCAAGCGGAGATGAAACCGGAGCTN 1380
 Db 1297 GCGAAGCTCAATTCAGAGGCTCACGAGATGCTCAAGCGGAGATGAAACCGGAGCTC 1356
 Qy 1381 CMTCTGCTCTGCTGCGGAGACCCNTCGCTCTCTATGATCTGCAAGGAGCTTCGATC 1440
 Db 1357 CCTCTCTCTCTGCGGCGGAGAGACCCCTGCTCTCTACCACTGCAAGGAGCTTCGATC 1416
 Qy 1441 GCGCGGAGGAGTCACTGCGAGGAGCTGAGGAGCTTCAAGCGGAGGAGGAGGAGGAG 1500
 Db 1417 GCGGCTGCGGAGTCACTGCGAGGAGCTGAGGAGCTTCAAGCGGAGGAGGAGGAG 1476
 Qy 1501 CAGCGGAGGAGTGGGCAACGAGGAGTCAAGGAGTCAAGGAGGAGGAGGAGGAGGAG 1560
 Db 1477 CAGCGGAGTGGAGTGGCAACGAGGAGTCAAGGAGTCAAGGAGGAGGAGGAGGAG 1536
 Qy 1561 ACGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 Db 1537 ACGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
 Qy 1621 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 Db 1597 CAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
 Qy 1681 NCGNCGTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 Db 1657 GTCGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
 Qy 1741 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 Db 1717 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
 Qy 1801 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 Db 1777 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
 Qy 1861 NNGTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
 Db 1837 TCGTCT 1885
 Qy 1921 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
 Db 1886 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1945
 Qy 1981 CGTCT 2040
 Db 1946 CGACT 2005
 Qy 2041 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
 Db 2006 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2065
 Qy 2101 TGAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
 Db 2066 TGAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2125
 Qy 2161 ACGTCT 2186
 Db 2126 ACGTCT 2151

RESULT 9
 US-10-374-366-9
 ; Sequence 9, Application US/10374366
 ; Publication No. US20040014085A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-song

APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/350,279
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
US-10-374-366-9

Query Match 65.6%; Score 1624.8; DB 16; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1828; Conservative 0; Mismatches 323; Indels 35; Gaps 5;

QY 1 ATGACCCCTCCCTCGACTCGATCGGACCTCGNTGCCAACGGCCTCCNAACGGATNG 60
DB 1 ATGACCCCTCCCTCGACTCGATCGGACCTCGNTGCCAACGGCCTCCNAACGGATNG 60
QY 61 CAGCCGNNCCGNNNGNCGNACGAGGCGCCAGTCCACNCTGNGNCGCGCGNG 120
DB 61 CAGCCGCT-----CAATGGGCGCTCGACCACTC-----GCACTCGAG 100
QY 121 GCTGCTCTCCGACCCAGCCAGNNAGCAGCTCGACATCGTNGAGNAGATCTCGCG 180
DB 101 GCTGACACTGCCACCAACCCAGTCAAGGAGTCACTGTCGAAGAAATGCTCGCG 160
QY 181 ACCCCACCCGNNAGNACGNNNTGAACTCGAGGGTCAACCTCACTCGAGAGCTCG 240
DB 161 CGCCGACCGACTGGA---CGCTGAACTCGAGCGGCTCACTCAACCTCGAGAGCTCG 217
QY 241 TCGGCGCCGCGCCGACGAGGCGCGCGCTCGGCTCGNACAGAGNAGAGATCGGCGCA 300
DB 218 TCTCGCGCGGAGGAGGAGGAGGCTGCTCGGCTCAAGAGACG-CAAGAGATCGCTCA 276
QY 301 AAGATCGACAAAGNGTCAAGTTCTCCGNNCNCAGCTTCNAACAAGNGTCTACGNGTC 360
DB 277 AAGATTGACAAATCGATCGAGTTCTTCGCGCTCGCAACTCTCCATAGAGGTCTACGCGCTC 336
QY 361 AACACTGGTTGGCGCGCTCGCGCGACCGCGGACTGAGAGAGTCAATTCCTCCAGAG 420
DB 337 AACACTGGATTTGGCGAGTCCGAGACACCGCGACCGAGAGGCAATCTCTCCAGAG 396
QY 421 GCNCTCTCGAGCAGCAGCTCGGCTGTCTCCNAGCTCGNTGANTCTTCNGCTC 480
DB 397 GCTCTCTCGAGCAGCAGCTCGGCTGTCTCCNAGCTCGNTGANTCTTCNGCTC 456
QY 481 GAGCGGCGCTCGAGAACTCGTTCCGCTCGAGGTCTCCGCGGCGCATCAATCCGCG 540
DB 457 GAGCGGCGCTCGAGAACTCGTTCCGCTCGAGGTCTCCGCGGCGCATCAATCCGCG 516
QY 541 GTCAACTCGCTCAACGCGCGCACTCGGCGGCTCGGCTCGTCTCGAGGCGCTCAAC 600
DB 517 GTCAACTCGCTCAACGCGCGCACTCGGCGGCTCGGCTCGTCTCGAGGCGCTCAAC 576
QY 601 AACTTCTCAACCAAGGATCAACCCCATCTGCTCCCTCGGCGGAGCATCTCGGCTCG 660
DB 577 AACTTCTCAACCAAGGATCAACCCCATCTGCTCCCTCGGCGGAGCATCTCGGCTCG 636
QY 661 GAGGAGCTCTCCGCTCTGNTACATCGCGCGCGGATCAACCGGTCAACCGGAGCANNAG 720
DB 637 GAGGAGCTCTCTCTCTCTCAATGAGCGCGCATCAACCGGTCAACCGGAGCANNAG 696
QY 721 GTNACGTTNNTCACGAGGAGCANNAGAAATCATGTNCGCCGCGAGGAGCATCGGCTC 780
DB 697 GTNACGTTNNTCACGAGGAGCANNAGAAATCATGTNCGCCGCGAGGAGCATCGGCTC 756
QY 781 TTNGGCTCGAGCGCGCTCTCTCGGCGCGGAGGAGGCTCTGCTCTGTCACGAGGAG 840

DB 757 TTCACTCGAGCGCGCTCTCTCGGCGCGGAGGAGGCTCTGCTCTGTCACGAGGAGC 816
QY 841 GCGCTTCGCGCTTCATGAGGAGCCTGCTCTGACAGAGCAGACAGTCTCTGCTCTC 900
DB 817 GCGCTTCGAGATGAGGAGCCTGCTCTGACAGAGCAGACAGTCTCTGCTCTC 876
QY 901 TCGAGCGCTCAAGGCTCTGACGCTGAGGAGCAGTGTGAGGAGGAGGAGGAGGAGGAG 960
DB 877 TCGAGCTCTCAAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
QY 961 CCNTTCTCAAGAGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 937 CCNTTCTCAAGAGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
QY 1021 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
DB 997 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
QY 1081 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
DB 1057 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
QY 1141 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB 1117 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGCAGCAGCAACCCGCTCATGAGCTCGAGAAAGAGAGCAGCAGGAGGAGCAG 1260
DB 1177 TCGAGCAGCAGCAACCCGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
QY 1261 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1237 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
QY 1321 GAGCAGCTCAACTTCAAGCAGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
DB 1297 GAGCAGCTCAACTTCAAGCAGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
QY 1381 CCNTCTCTCTGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
DB 1357 CCNTCTCTCTGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
DB 1417 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
QY 1501 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
DB 1477 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
QY 1561 AAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
DB 1537 AAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
QY 1621 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
DB 1597 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656
QY 1681 NCNNGCTCANTNAGCAGCACTTTGAGCNGCCTCGACCGGNNNNNNNNNNNNAGAA 1740
DB 1657 GTCGCTCATGACAGCACTTTGAGCNGCCTCGACCGGNNNNNNNNNNNNAGAA 1716
QY 1741 CTGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
DB 1717 CTGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
QY 1801 CTGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
DB 1777 CTGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
QY 1861 NNGTCTCTGCGGNNNNGCANNAGGCTCTGCTCGGAGGAGGAGGAGGAGGAGGAGGAG 1920
DB 1837 TCGTCAAGCT-----GCTCTGCTCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1885

QY 1921 CGCCGAGAGGACATCTTCGTCTACGCGCGANANGTCGCGACNCCTTTCTGAGCGCNCCT 1980
| | | | |
Db 1886 CGCCCAAGTCGGCATCTCTGCTACCCGCCAAGTCGCGAGACCTTCTGGGCCCGCGT 1945
| | | | |
QY 1981 CGTCGTCTGCGCGCGGCTCNCGTACTCTCTGCGCGGACGCGGCTCTGACTGTTG 2040
| | | | |
Db 1946 CGACCTCTGCGCGCGGCTCTCTGTACTCTCTGCGCGGACCTCAATCTCTTACGCCCTTCG 2005
| | | | |
QY 2041 TCCGCGAGAGCTCGCGCTCAAGGCCCCGCGCGGACGCTTCTCTCGGCAAGAGAG 2100
| | | | |
Db 2006 TCCGCGAGAGCTTGGCGCTCAAGGCCCGCGCGAGACGCTCTTCTCGGCAAGAGAG 2085
| | | | |
QY 2101 TGAAGATCGGACCAAGCTCTCCGATCTTACAGAGCCATCAAGNCGGCGCATCAAC 2160
| | | | |
Db 2066 TGAAGATCGGACCAAGCTCTTCAAGATCTACAGGCGCATCAAGTCGGGCGAGATCAAC 2125
| | | | |
QY 2161 AGCTCTCTGTCAGAGATGCTCGCNTAG 2186
| | | | |
Db 2126 AGCTCTCTCTCAAGATGCTCGCTTAG 2151
| | | | |

RESULT 10
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; PRIOR FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-13

Query Match 65.6%; Score 1624.8; DB 16; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1828; Conservative 0; Mismatches 323; Indels 35; Gaps 5;

QY 1 ATGCGCCCTCCNTGACTCGATCGGACCTCGACCTCGANTCGCAAGGCGTNCMAACGANTNG 60
| | | | |
Db 1 ATGCGACCCCTCGCTGACTCGATCTTGCACTGTTGCAAAAGGCGTGCATCCGCAAG 60
| | | | |
QY 61 CAGCGCCGNNCGCNCNNNGCNCNACGCGCGCCAGCTCCACNCTCNGCNGCGCNG 120
| | | | |
Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GAGTGGAG 100
| | | | |
QY 121 GCTCGCTCTCTCCGACACCAACCCAGNNAGACCTGACATCTGTNAGNAGATCTCTGCG 180
| | | | |
Db 101 GCTCGACCTCTGCGCAACCAACGAGTCAAGAGTGCATCTGTCAAGATGCTTGGCG 160
| | | | |
QY 181 ACCCGACCGNNAAGNAGNNNTGAACTCGAGGAGTACACCTTCACTCGAGNAGCTCG 240
| | | | |
Db 161 CGCGCGACCACTGGA---CGCTGAACTCGAGGCTACTCGTCAACTCGGAAGAGCTCG 217
| | | | |
QY 241 TCGGCGCGCGCGCAAGGCGCGCNCNGTCCGCTGNCAGACAGNCGAGAGATCCGCGCA 300
| | | | |
Db 218 TCTCGCGCGCGAGAGAGGCGAGGCTGTCCGCGTCAAGAGACAG--GAGAGAGATCGCTCA 276
| | | | |
QY 301 AAGATGACAAANAGNCTGATTCCTCGGNNCNCAGCTGNAACAAGAGTCTACGAGNCTC 360
| | | | |
Db 277 AAGATTGACAAATCGGTGAATTTCTTGCGCTGCAACTCTCATAGAGGCTTACGAGCTC 336
| | | | |
QY 361 ACGACTGTTTCGCGCGCTCGCGCGACACCGAGTGAAGATGNAATCGCTCCAGAG 420
| | | | |

Db 337 ACGACTGATTTGGGGATTCGCGACACCCGCAACCGAGACGCGCATCTCGCTCCAGAG 336
| | | | |
QY 421 GNCCTCTTGAGACACACCTCTGCGGTGTNCTCCNAGTTCGANTCGANTCCTTNGCCTC 480
| | | | |
Db 397 GCTCTCTTGAGACACACCTCTGCGGTGTNCTCCNAGTTCGANTCGANTCCTTNGCCTC 456
| | | | |
QY 481 GNNCGCGGCTCGAAGATCTGCTTCCGCTCGAGGTGCGCGCGGCGGCGCATACCATCCG 540
| | | | |
Db 457 GCGCGGCTCTGAGAAATCTGCTTCCCTCGAGGTGTTCGCGGCGGCGCATACCATCCG 516
| | | | |
QY 541 GTCAACTGCTCAACNCGGCGCACTGCGCNGTCCGCTCGTGTCTTGAGGCGCTCAC 600
| | | | |
Db 517 GTCAACAGTTTACCCGCGGCACTCGGCTGTCCGCTCGTGTCTTGAGGCGCTCAC 576
| | | | |
QY 601 AACTTCCTCAACAGAGATACCCCGATCGTCCCGCTCGGCGGCGCAATCTCGGCGT 660
| | | | |
Db 577 AACTTCCTCAACAGAGATACCCCGATCGTCCCGCTCGGCGGCGCAATCTCGGCGT 636
| | | | |
QY 661 GCGGACTCTCCCGCTCTGNTACATCGCGCGCGCATCAACGCTCAACCGGACNNAG 720
| | | | |
Db 637 GCGGACT 696
| | | | |
QY 721 GTNCAAGTNTNCAAGAGGCGCANAGAAATCATGTNCGCCCGCGAGGCGATCGGCTC 780
| | | | |
Db 697 GTGCAAGTGTTCAGAGAGGCGCAAGAGAAATCTGTACGCGCGAGGCGATGCGCTC 756
| | | | |
QY 781 TTNGCTCTGAGCCCGTCTGCTCTCGGCGGAGAGGCTCGGTCTCGTCAACGAGCAG 840
| | | | |
Db 757 TTCAACTCTGAGCCCGTCTGCTCTCGGCGGAGAGGCTCTCGGTCTCTGTCACGCGAC 816
| | | | |
QY 841 GCGGTCTCGGCTGATGAGGAGCCCTGCTTGCACAGCGCACATGCTCTCGCTCTC 900
| | | | |
Db 817 GCGGTCTGAGCATGATGAGGAGCCCTGCTTGCACAGCGCACATGCTCTCGCTCTC 876
| | | | |
QY 901 TCGGAGGCGCTCAGGCGCTTNAAGGTGAGGCGCATGTGTGCGGCGCGCGGCTCGTTCCAC 960
| | | | |
Db 877 TCGAGGTGCTTACAGGCGCATGAGGTCAAGCGATGTGTGCGGCGCGCGGCTCGTTCCAC 936
| | | | |
QY 961 CANTTCTCTCAAGAGTCAAGCGCCCTTACCGGACCGCATGATGAGGTGCGGCAATC 1020
| | | | |
Db 937 CCTTCTCTTCAAGAGTCAAGCGGCGCTTACCGGAGCGAGTGAAGTTCGCGGAAACATC 996
| | | | |
QY 1021 CGCAGCTCTCTGAGAGGAGAGNNTTTCGCTTCCATCCACAGAGAGGAGGTCAAGTCAAG 1080
| | | | |
Db 997 CGCAGCTCTCTGAGAGGAGGCGGCTTGTCTTCCATCCATGAGAGAGGTCAAGTCAAG 1056
| | | | |
QY 1081 GACGAGAGGCGATTTCTCGGCGAGGAGCGCTACCGGCTCCGCAAGTGCAGTGGCTC 1140
| | | | |
Db 1057 GAGAGAGAGGCGATTTCTCGGCGAGGAGCGGCTTACCGGCTTGGCGAGTCTCTCAAGTGGCTC 1116
| | | | |
QY 1141 GGCCTGCTGTGAGGAGATGATTTACGCGCCACCGGCTCTCTGCTGAGGCGGNCAG 1200
| | | | |
Db 1117 GGCCTGCTGTGAGGAGATGATTTACGCGCCACCGGCTCTCTGAGGCGGCGCAG 1176
| | | | |
QY 1201 TCGAGAGCGGACCAACCGGCTATCGAGTTCAGAGAAACAAGNAGACCAACGCGGCGCAAC 1260
| | | | |
Db 1177 TCGAGAGCGGACCAACCGCTTCTATCGAGTTCAGAGAAACAAGTTCGAGCGGCGGCAAT 1236
| | | | |
QY 1261 TTTCAGAGCGNCGCTGTGCGNNAACGATGAGAAAGACTGCTTGCNCTGCGCTGATC 1320
| | | | |
Db 1237 TTTCAGAGGCTGCGGCTGTGCGCAACCAATGAGAAAGACTGCTGCGCTGCGGCTGCGCAGATC 1296
| | | | |
QY 1321 GCGAGACTTAATTCAGAGAGTCAACCGAGATGCTCAACGCGCGGATGAACGCGGCGCTN 1380
| | | | |
Db 1297 GCGAGACTTAATTCAGAGAGTCAACCGAGATGCTCAACGCGCGGATGAACGCGGCGCTC 1356
| | | | |
QY 1381 CANTCTGCTCTGCTGCGAGAGACCCNTGCTCTCTCTATCACTGCAAGGCGCTCGACATT 1440
| | | | |
Db 1357 CCTCTGCTCTGCGGCGGAAAGACCCCTGCTCTCTCTCTCTCTGCACTGAGGCGCTCGACATC 1416
| | | | |
QY 1441 GCGCGNCGNCTAAGACTTTCGAGCTCGAGNCACTTGGCAACCGGATTAACCAACCACTC 1500
| | | | |
Db 1417 GCGCGTGGCGGCTAAGCTTCGAGGTTGGAGACCTTGGCAACCTTGGTGAAGACGATATC 1476
| | | | |

Qy	1501	CACCGCGNAGATGCGGCAACAGGCGCGTCACTCGCGCNCATCTCGGCGNCGCG	1566
Db	1477	CACCGCGCTGAGTGGCGAACCAGGCGGTCACTCGCGCTCACTCGCGTGGTGGC	1538
Qy	1561	ACNCGCGAGGCAACGACGTCCTTTCCTCTCTCTCTCGCCACACCGCTCTACTGCGTCTC	1620
Db	1537	ACGACCGAGTCCAGAGAGTCTTTCCTCTCTCTCTCTCGCCACACCGCTCTACTGCGTCTC	1596
Qy	1621	CAGGCGGTGACCTCCGGCGATGAGAGTTGCAATTCAGAGAGAGTTTCGACCCGATNNITC	1688
Db	1597	CAAGCCATGACCTTGGCGCGACCCAGTTTCAGTTTCAAGAGAGAGTTTCGCGCCACGATC	1656
Qy	1681	NCNNCCCTNTCAGAGAGACTTTGGCNCNCCTTCACAGCGCANNNNNNNNNNNNAAGAA	1748
Db	1657	GTCTCGCTATGACCAAGACTTTTGGTCCCGCATGACCGGCTCGAAGCTGGCGGACGAG	1716
Qy	1741	CTCGNGGACAAAGTCAACAAGCGCTCNAACAAGCNCCTCGAGGACAACTCTGTACAC	1800
Db	1717	CTCGTGAAAGATTAACAAGACGCTCGCCCAAGCGCTCGAGGACAACTCTGTACAC	1767
Qy	1801	CTCGAGCCGCGCTGGGACGAGCGCCTTCCTCGTTGGCGACGGGACCGATCGTCAGANNNNN	1860
Db	1777	CTCGTCCCGCGCTGGACGAGCGCTTCTCTTTCGCGCGCGGACCGCTGTCAAGTCTCTC	1836
Qy	1861	NNGTCTCGCCNNNNNGCCANNAGTCTCGCTCGCNCGCGTCAACGCGCTGAAAGTTCGCT	1920
Db	1837	TCGTCAACGTC-----GCTCTCGCTCGCGCGCTCGCAACGCTGGAAGTTCGCGG	1885
Qy	1921	CGCGCGAGAGGCGCATCTCGCTCAACGCGGNNAGTCCGCGACNCCTTCGAGNCGCNCGT	1980
Db	1886	CGCGCGAGTCGCGCATCTCGCTCAACGCGCGCAAGTCCGCGGACCTTCTGTGTCGCGCGCT	1945
Qy	1981	CGTGTGCTGCGCGCGCTCNCGTACCTTCGTCGCGGACGCGCGTCTGTACTGTCGTTG	2040
Db	1946	CGACCTCGTGGCGCGCTCTGTACCTCTGGCGCGCACTCAGATCTCTAGCGCTTGG	2005
Qy	2041	TTCGCGAGAGACTCGGCGCTCAAGGCGCGCGCGCGGCGACGCTTCTCTCGGCAAGCGAGG	2100
Db	2006	TTCGCGAGAGGCTTGGCGTCAAGGCGCGCGCGGAGCGCTTCTCTCGGCAAGGAGG	2065
Qy	2101	TGACGATCGGCAACAAGCTTCGCGCATCTACGAGGCGATCAAGNNCGGCNCATCAAC	2160
Db	2066	TGACGATCGGCTGAAAGCTTCAAGATCTACGAGGCGCATCAAGTGGGCGAGATCAACA	2125
Qy	2161	AGGTCTGTCAAGATGCTCGCNTAG	2186
Db	2126	AGGTCTCTCAAGATGCTCGCNTAG	2151

```

RESULT 11
US-10-374-366-23
? Sequence 23, Application US/10374366
? Publication No. US20040014085A1
? GENERAL INFORMATION:
? APPLICANT: Tang, Xiao-Song
? APPLICANT: Milano, Joseph
? TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
? FILE REFERENCE: C11794 US NA
? CURRENT APPLICATION NUMBER: US/10/374,366
? PRIOR FILING DATE: 2003-02-26
? PRIOR FILING DATE: 2002-02-26
? NUMBER OF SEQ ID NOS: 203
? SOFTWARE: Patencin version 3.2
? SEQ ID NO 23
? LENGTH: 2151
? TYPE: DNA
? ORGANISM: Rhodospiridium rubrum
US-10-374-366-23

```

Query Match	DB 16;	Length 2151;
65.6%;	Score 1624.8;	
Best Local Similarity	83.6%;	Pred. No. 0;

Matches 1828; Conservative 0; Mismatches 323; Indels 35; Gaps 5;

QY	ATGAGCCCTCCNTGACATCGACATCGCACCTCGTGGCCGACAGGAGNTGCMNACAGGNTG	60
Db	1 ATGAGCACTTCGTGACTCGATCTTCGCACTCGTTGGCAAAAGGCGTCCGATCCGAAAG	60
QY	61 CAGCGCAGNCCGNCNNNGNCNACAGGGCGCCAGTCGCACNTGNCNGGCCGCGNG	120
Db	61 CAGGCTGT-----CAATGGCGCCTCGACCAACTCT-----GGAGTCGAG	100
QY	121 GCTGCTCTCTCCCGACCAACCGAGNNNGAGAGCTCGCATCTGNNNGNAGATCTCTCGCG	180
Db	101 GCTGCACTTGCCCAAAACCGAGTCAGGAGGTGACATGCTCGGAAAGATGCTCGCG	160
QY	181 ACCCCACCGNNAAGNACGANNNTGAACTCGAGGGGTACACCTCACCTCGAGNAGCTCG	240
Db	161 CGCGGACCGACTCGA---CGCTCGAACTCGACGGCTACTGCTCAACTCGGAGACGTG	217
QY	241 TGGGGCGCGNCNGCAAGGGCGCNGCNGTCGCGGNCAGACAAGNAGAGATTCGAGCA	300
Db	218 TCTCGCGCGCGAGAAAGGCAAGGCTGTGCGCTCAAGACAG-CAGACAGATCCGCTCA	276
QY	301 AAGATGACAAANAGNGTCGAGTTCCTCCGNNCAGCTCNACAAAGNGTCTACGAGTTC	360
Db	277 AAGATTGACAAATCGGTGAGATTCTTGCGCTCGCAACTCCATGAGCCTCTACGGCGCTC	336
QY	361 AGACATGATTGCGGGGCTCGGCGCGACACCGGACTGAGAGATGNCATCTCGTCCAGAG	420
Db	337 AGACATGATTGGCGGATTCGCGAGAACCCGACCGAGGAGCGCATCTCGCTCCAGAG	396
QY	421 GNCCTCTTGAGACACCACTGTGGGGTGTNCTCCNACGTCGNTGANTCTTNGCCTC	480
Db	397 GCTCTCTCGAGACACCACTGTGGGGTGTCTCCCTGCTGTTGCACTCGTTCGCGCTC	456
QY	481 GAGCGCGGCTCGAATCTGCTTCGCTCGAGGTGTGCGGGGCGCATGACATTCGCG	540
Db	457 GCGCGGGTCTGAGAACTCGCTTCCCTCGAGGTGTGTGGGGCGCATGACATTCGCG	516
QY	541 GTCACTGCTCANCNGCGGCGCACTCGGCGNGTCCGCTCGTGTCTCGAGGCGCTCAC	600
Db	517 GTCAACAGTTGACCCGCGGCACTCGGCTGTCCGCTGTGCTCTCGAGGCGCTCAC	576
QY	601 AACTTCTCAACCAAGGATCAACCCCATGTCCTCCCTCGGGGACCAATCTCGCGCTCG	660
Db	577 AACTTCTCAACCAAGGATCAACCCCATGTCTCCCTCGGGGACCAATCTCTGCTGCG	636
QY	661 GCGGACTTCCCCNGCTCTCNATCATGCGCGCGCATCAACGGTCAACCCGAGCNCNAG	720
Db	637 GCGGACTTCTCTCTCTCTCTCAATGACAGCGGCATCAAGGAGTCAACCCGAGCAGAG	696
QY	721 GTNCAAGTNNNCAAGAGGCGANNAGAAATCATGTNNGCCGCGAGGCGATCGCGTTC	780
Db	697 GTGCAAGTGTGCAAGAGGCGAAGAAATCTGTAGCCCGCGAGGCGATGCGCGCTC	756
QY	781 TTNGTCTTGAAGCCCGTGTCTTGCGCCGAGAGAGGTTCTGCTCTGTCACAGGACG	840
Db	757 TTCAACCTGAGCCCGTGTCTGTGCGCCGAGAAAGAGGTCTCGGTCTGTCACAGGAC	816
QY	841 GCGGTCTCGGCTCGATGGGAGCCCTGCTCTGCAAGAGACACATGTCTCGTCTCTC	900
Db	817 GCGGTCTGAGATCGATGGCGACCTTGTCTGTGACAGAGCACACATGTCTCGGCTCTC	876
QY	901 TCGAGAGGCTCAAGGCTCTNACGCTGAGAGCGATGTCGCGCACGCGCGGCTGTTCCAC	960
Db	877 TCGAGAGGCTCAAGGCGCATGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	936
QY	961 CCNTTCTCTCAAGAGTCAAGCGGCTCTCAACCGACCCATGAGAGTTCGCGCGCAATC	1020
Db	937 CCTTCTCTCAAGAGTCAAGCGGCTCTCAACCGACCCATGAGAGTTCGCGGAAATC	996
QY	1021 CGAGGCTCTCGAGAGGCGAGGNTTGGCGTCTCAACCAAGAGAGAGTCAAGGTCAAG	1080
Db	997 CGCAGGCTCTCGAGAGGCGAGGNTTGGCGTCTCAACCAAGAGAGAGTCAAGGTCAAG	1056

QY	2161	ACGTCCTCGTCAAGATGCTGCNTAG	2186
Db	2126	ACGTCCTCTCTCAAGATGCTGCTTAG	2151
RESULT 12			
	US-09-765-873A-7		
	; Sequence 7, Application US/09765873A		
	; Patent No. US20010053847A1		
	GENERAL INFORMATION:		
	; APPLICANT: Tang, Xia-Song		
	; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID		
	; FILE REFERENCE: BC1009 US CIP		
	; CURRENT APPLICATION NUMBER: US/09/765,873A		
	; CURRENT FILING DATE: 2001-01-19		
	; PRIOR APPLICATION NUMBER: US 09/627,216		
	; PRIOR FILING DATE: 2000-07-27		
	; PRIOR APPLICATION NUMBER: US 60/147,719		
	; PRIOR FILING DATE: 1999-08-06		
	; NUMBER OF SEQ ID NOS: 38		
	; SOFTWARE: Microsoft Office 97		
	; SEQ ID NO 7		
	; LENGTH: 2151		
	; TYPE: DNA		
	; ORGANISM: Rhodotorula glutinis		
	; FEATURE:		
	; NAME/KEY: CDS		
	; LOCATION: (1)..(2151)		
	; US-09-765-873A-7		
Query Match	65.6%;	Score 162.2;	DB 9; Length 2151;
Best Local Similarity	83.6%;	Pred. No. 0;	
Matches 1827;	Conservative	0;	Mismatches 324; Indels 35; Gaps 5;
QY	1	ATGGGCCCCCTCCMTCCGATCGATGGAGGAGCTCGTGGCCAAAGGCGTCTCNAAAGGANTNG	60
Db	1	ATGGGACCCCTCGCTCGATCGATTCGCACTCGTTTGCAAAAGGCGTCCGATCCGCAAG	60
QY	61	CACGCGCNCNCGNCNNNGCNCNACGAGGCGCCAGCGTCCACNCTCGNCGCGCNCNG	120
Db	61	CAGGCTGT-----CAATGCGGCTCGACCAACCTC-----GCAGTCGAG	100
QY	121	GCTGCTCTCTCCCGACCAACCAGAGNAGCGAGCTCGACATCTGTNAGNAGATCTCTCGCG	180
Db	101	GCTGCACTCTCCCAACAACCGATCTACGAGGTGACATCTGTGAGAAGATGTTGCGCG	160
QY	181	ACCCACCGCAGNACGNAAGNNCTGAACTCGACGCGGTACACCTCCACCTCGGAGCAGTCG	240
Db	161	CGCCGACCGAATCGA---CGGTGAACTGACGGGTAACTCGTCAACCTCGAGACGTCG	217
QY	241	TGGGCGCCGCGCNCGCAAGGCGCGCNCNGTCCGCGTCAAGACAGNACGAGATCCGCGCA	300
Db	218	TCTCGCGCGCAGAGAAAGGCGCAGGCGTGTCTCGCGTCAAGACAG-CGACAGATCCGCTCA	276
QY	301	AAGATCGACANAAAGNGTCGAGTTCTCTCCGNNCAGACTGNACAAAGNATCTAAGAGNCTC	360
Db	277	AAGATTGCAAAATCGGTTCGAGTTCTTGCGCTCGCACTCTCAATGAGCGTTACGCGCTC	336
QY	361	ACGACTGATTTGGCGGCGCTCGCGCGACACCCGAGCTGAGAGTGCNATCTCGTCCAGAA	420
Db	337	ACGACTGATTTGGCGGATCCGCGAGACACCCGACCGAGAGAGCGCCATCTCGCTCAGAA	396
QY	421	GCGTCTCTCGAGACACAGATCTTGCGGATGTCCTCCNAGCGTCAATCTTTCNGGCTTC	480
Db	397	GCTCTCTCGAGACACAGATCTTGCGGATGTCCTCCNAGCGTCAATCTTTCNGGCTTC	456
QY	481	GGNCGCGGCTCGAGAACTCGATTTCGCTGAGAGTGTCTCGGCGCGCATGACATCCGCG	540
Db	457	GCGCGCGGCTCTCGAGAACTCGATTTCCTCCCTGAGAGTTGTTGCGGCGCATGACATCCGCG	516
QY	541	GTCGAATCTGCTCAACGCGGCGCACTGGGCGNCTCGCTGTGTCTCTGAGGCGCTCAAC	600
Db	517	GTCGAACACTGTAACCGCGCGCACTGGGCGNCTCGCTGTGTCTCTGAGGCGCTCAAC	576

QY 601 AACTTCCTCAACCAAGGAGATCAACCCCATCTGTCCTCGGAGGACCATCTCGAGCTCG 660
 Db 577 AACTTCCTCAACCAAGGAGATCAACCCCATCTGTCCTCGGAGGACCATCTCGAGCTCG 636
 QY 661 GAGGAGCTCTCCCT 720
 Db 637 GAGGAGCT 696
 QY 721 GTTACAGTNN 780
 Db 697 GTTACAGTNN 756
 QY 781 TTNGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 Db 757 TTGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816
 QY 841 GCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
 Db 817 GCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876
 QY 901 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 Db 877 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
 QY 961 CATTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 Db 937 CATTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
 QY 1021 CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 Db 997 CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
 QY 1081 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 Db 1057 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
 QY 1141 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 Db 1117 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
 QY 1201 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 Db 1177 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
 QY 1261 TTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 Db 1237 TTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
 QY 1321 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 Db 1297 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
 QY 1381 CATTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 Db 1357 CATTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
 QY 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 Db 1417 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
 QY 1501 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 Db 1477 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
 QY 1561 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 Db 1537 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
 QY 1621 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 Db 1597 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1656

QY 1681 NCNNGCTCCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 Db 1657 GTTCTGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
 QY 1741 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 Db 1717 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
 QY 1801 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 Db 1777 CTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
 QY 1861 NGTCTGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
 Db 1837 NGTCTGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1885
 QY 1921 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
 Db 1886 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1945
 QY 1981 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
 Db 1946 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2005
 QY 2041 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
 Db 2006 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2065
 QY 2101 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
 Db 2066 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2125
 QY 2161 ACCTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2216
 Db 2126 ACCTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2181

RESULT 13
 US-10-188-523B-7
 ; Sequence 7, Application US/10188523B
 ; Publication No. US20030170834A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-Song
 ; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
 ; FILE REFERENCE: BCI009 US DIVCIP
 ; CURRENT APPLICATION NUMBER: US/10/188,523B
 ; PRIOR FILING DATE: 2002-10-17
 ; PRIOR APPLICATION NUMBER: US 09/627,216
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR APPLICATION NUMBER: US 60/147,719
 ; PRIOR FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 7
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: Rhodotorula glutinis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2151)
 ; OTHER INFORMATION:
 ; US-10-188-523B-7

Query Match 65.6%; Score 1623.2; DB 15; Length 2151;
 Best Local Similarity 83.6%; Pred. No. 0;
 Matches 1827; Conservative 0; Mismatches 324; Indels 35; Gaps 5;
 QY 1 ATGCCCCCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
 Db 1 ATGCCCCCTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
 QY 61 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
 Db 61 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120

Db 61 CAGGCTGT-----CAATGCGCTTGACCAACCTC-----GCACTCCGAG 100
QY 121 GCTGCTCTCTCCGACCAACCGANNNGACGACGTGACATGTTNGAGNAGATCTCTCCGC 180
Db 101 GCTCGACCTGCCCCAACCCAGGTCAAGCAGGTGACATGCTGAGAGAGATGCTCCGCG 160
QY 181 ACCCGACCGNNMGACGNNMTGAACTCGAGGGGTACACCTCACTCCGNNACGTGCG 240
Db 161 CGCCGACCGACTGGA---CGCTGAACTGAGGGCTACTGCTCAACTCGAGACGTGCG 217
QY 241 TCGGCGCCGCGNCGAAGGCGCGCNGCTCCGCTGACAGACGACGACGAGATCCGCGCA 300
Db 218 TCTCGGCGCGGAGAGGCGAGGCGCTGCTCCGCTGAAAGACAG--GAGAGAGATCCGCTCA 276
QY 301 AAGATCGAATAAAGTGTGAGTTCCTCCGNNCNGCTCTMAACAAGTGTCTAAGGNTGTC 360
Db 277 AAGATTGAAATAATCGGTGAGTTCCTGCGCTCGCACTCTCCATGAGGCTGTACGAGGTC 336
QY 361 ACGACTGTTTTGCGCGGCTCGGCGGACCCGGAAGTGAATGATGCTGCTCCAGAG 420
Db 337 ACGACTGGAATTTGGGGGATCCGCGACACCGGACCGAGAGCGCATCTGCTCCAGAG 396
QY 421 GCNCTCTCGAGACCAACGCTCTGCGGTGTTCCCNACGTGANTGANTCTTCGCTC 480
Db 397 GCTCTCTCGACACACAGCTCGCGGTGCTCTCCCTTCTGCTGCTGCACTGCTCCGCTC 456
QY 481 GGNCGGCGCTCGAGAACTGCTTCGCTCGAGGTGCTCGGCGGCGGCGCATGACATCGCG 540
Db 457 GCGCGGCGCTCGAGAACTCGCTTCCTCGAGGTGTTTCGGGCGCGCATGACATCGCG 516
QY 541 GTCAAATCTGCTCAACNCGCGGCGCACTCGGCGTCTGCTGCTCGTCTGAGGCGCTCAC 600
Db 517 GTCAACAGTTGACCCGCGGCGCACTCGGCTGCTCGCTGCTGCTGCTGAGGCGCTCAC 576
QY 601 AACTTCTCTAACAACAGGCGATCAACCCCATCTGCTCCCTCCGCGGACCATCTCGGCTCG 660
Db 577 AACTTCTCTAACAACAGGCGATCAACCCCATCTGCTCCCTCCGCGGACCATCTCGGCTCG 636
QY 661 GCGGACCTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 637 GCGGACCT 686
QY 721 GTNCAAGTNNNTACAGAGGCGANNAGAAATCATGTNCGCGCGGAGCGGATCGCGCTC 780
Db 697 GTGCAAGTGTGTCAAGAGGCGAAGAGAAATCTGTAGCGCCCGAGAGCGATGCGCTC 756
QY 781 TTNGGCTCGAGCCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 757 TTCAAACCTGAGCCCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
QY 841 GCGCTCTCGGCTCGATGCGGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 817 GCGGCTCGAGCATGATGAGGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
QY 901 TCGCAGAGCGCTCAAGGCTCTTACGATGAGGCGATGCTGCGGCGCGGCTGCTTCCAC 960
Db 877 TCGCAGTGTGCTCAAGGCGATGAGCGGTTCMAAGATGCTGCGGCGCGGCTGCTTCCAC 936
QY 961 CCNTTCTCTCAAGAGTCAAGCGGCTTCAAGCCGAGCCGATGAGGTGCGGCGCAATC 1020
Db 937 CCTTCTCTTCAAGAGTCAAGCGGCTTCAAGCCGAGCCGATGAGGTGCGGCGCAATC 986
QY 1021 CGCAGGCTCTCTGAGGCGCGNNGTTTGGCTTCCGCTCAACGAGAGAGAGTCAAGTCAAG 1080
Db 997 CCGAAGCTCTCTGAGGGAAGCGCTTGTCTTCCACATGAGGAGAGTCAAGTCAAG 1056
QY 1081 GAGCAGAGAGGAGATCTCTCGCGCAGAGACGCTACCGGCTCCGCGCATGCTGCTGCTC 1140
Db 1057 GAGCAGAGAGGAGATCTCTCGCGCAGAGACGCTACCGGCTCCGCGCATGCTGCTGCTC 1116
QY 1141 GCGCGGCTCTGAGGAGATGATTCACGCGCCACGCGGCTCTCTGCTGCTGAGGCGGAG 1200
Db 1117 GCGCGGCTCTGAGGAGATGATTCACGCGCCACGCGGCTCTCTGCTGCTGAGGCGGCGAG 1176

QY 1201 TCGACACCGACCAACCGGCTCATGACGCTGAGAAACAAGANAGACCAACGAGGCGAAC 1260
Db 1177 TCGACACCGACCAACCGGCTCATGACGCTGAGAAACAAGATTGCGACCAACGAGGCGAAC 1236
QY 1261 TTCGAGGCGCGGCTGCTGCGNAAACAGATGAGAGAGACTGCGGCTGCGNCTGCGCTGATC 1320
Db 1237 TTCGAGGCTGCGGCTGCTGCGNAAACAGATGAGAGAGACTGCGGCTGCGGCTGCGGCTGATC 1296
QY 1321 GCGAAGCTCAATTCAACGACGCTCAACGAGATGCTCAACGCGGCGATGAACCGGCGCTTN 1380
Db 1297 GCGAAGCTCAATTCAACGACGCTCAACGAGATGCTCAACGCGGCGATGAACCGGCGCTTN 1356
QY 1381 CCCTCTCTCTGCGGCGGAGAGACCCCTGCTCTCTCAACGAGAGGCTCTGAGATC 1440
Db 1357 CCCTCTCTCTGCGGCGGAGAGACCCCTGCTCTCTCTCAACGAGAGGCTCTGAGATC 1416
QY 1441 GCGGCGGCGGCTCACTTCCGAGCTCGAGACCTTCCGACCCGCTTCAACGAGAGGCTCTGAGATC 1500
Db 1417 GCGGCTGCGGCGGCTCACTTCCGAGCTCGAGACCTTCCGACCCGCTTCAACGAGAGGCTCTGAGATC 1476
QY 1501 CAGCGCGGCGGAGATGAGGCAACGAGCGGCTCAACTGCTGCGNCTCACTCTGAGGCGCGC 1560
Db 1477 CAGCGCGGCTGAGATGAGGCAACGAGCGGCTCAACTGCTGCGNCTCACTCTGAGGCGCGC 1536
QY 1561 ACGGCGAGGCGCAACGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1537 ACGGCGAGGCTCAACGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGGCGGCTGAGCTCTGCGGCGGAGTGAAGTTCGAGTTCAAGAGAGCTTCAACCCGNTNCT 1680
Db 1597 CAGGCGGCTGAGCTCTGCGGCGGAGTGAAGTTCGAGTTCAAGAGAGCTTCAACCCGNTNCT 1648
QY 1681 NONNCGCTCTCAACGAGCTTGGGCGGAGCTTGGGCGGAGCTTGGGCGGAGCTTGGGCGGAG 1740
Db 1657 GTCTGCTCTATGACGAGCACTTGGGCGGAGCTTGGGCGGAGCTTGGGCGGAGCTTGGGCGGAG 1716
QY 1741 CTCGAGGCAAGAGTCAACAAGGCTGCAACAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAG 1800
Db 1717 CTCGAGGCAAGAGTCAACAAGGCTGCAACAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAG 1776
QY 1801 CTCGAGGCGGCTGAGGAGGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 1777 CTCGAGGCGGCTGAGGAGGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1836
QY 1861 NNGTCTGCGGNNNGCANNAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1837 TCGTGAAGCTC-----GCTTGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1885
QY 1921 CCGCGGAGAGGAGCTCTGCTCAAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 1886 CCGCGGAGGAGGAGCTCTGCTCAAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1945
QY 1981 CGTCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
Db 1946 CGAAGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2005
QY 2041 TCGCGGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Db 2006 TCGCGGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2065
QY 2101 TCGAGATGAGCAACGCTTCCGCGCATCTAAGGCGCATCAAGGCGGCGGCGGCGGCGGCGGCGG 2160
Db 2066 TCGAGATGAGCAACGCTTCCGCGCATCTAAGGCGCATCAAGGCGGCGGCGGCGGCGGCGGCGG 2125
QY 2161 ACGTCTCTGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2186
Db 2126 ACGTCTCTGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151

RESULT 14
US-10-374-366-1
; Sequence 1, Application US/10374366


```
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; PRIORITY FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIORITY FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodospiridium glutinis
; US-10-374-366-1

Query Match      65.6%; Score 1623.2; DB 16; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1827; Conservative 0; Mismatches 324; Indels 35; Gaps 5;

QY 1 ATGGCCCCCTCCCTGCACTCGATCGGACCTCGGCAAGCGGTCGCAAGCGTNG 60
DB 1 ATGGCACTCCCTCGCTCGACTCGATCTGCACTGTTGCAAGGGGTGCGATCCGCAAG 60
QY 61 CAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GCACTCGGAG 100
QY 121 GCTGCTCTCTCCGCAACCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 101 GCTGCGACCTCGCCCAACCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 160
QY 181 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 161 CGCGGCACTCGA---CGCTCGAAGTGAAGGCTACTCGCTCAACCTCGGAGACGTCG 217
QY 241 TCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 218 TCTCGGCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
QY 301 AAGATGCAAGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 277 AAGATTGCAAAATCGTCTGAGTCTTGCGCTGCAACTCTCCATGAGGCTTACGCGTC 336
QY 361 ACGACTGCTTTCGCGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
DB 337 ACGACTGATTTGGCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
QY 421 GCGCTCTCTGAGGAGCAAGCTTGGCGGCTGCTCCGAGCGTCAAGTCAAGTCT 480
DB 397 GCTCTCTCTGAGGAGCAAGCTTGGCGGCTGCTCCGAGCGTCAAGTCTGAGTCT 456
QY 481 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
DB 457 GCGCGGCGGCTCGAAGACTCGCTCGCTCGAGGCTGCTCGGCGGCGGCGGCGG 516
QY 541 GTCAACTGCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 517 GTCAAGCTTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
QY 601 AACTTCTCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
DB 577 AACTTCTCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
QY 661 GCGGCGGCTCTCCGCTCTCTCAATCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
DB 637 GCGGCGGCTCTCTCTCTCTCAATCGGCGGCGGCGGCGGCGGCGGCGGCGG 696
QY 721 GTTCACTTNTTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
DB 697 GTTCACTTNTTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
```

```
QY 781 TTNGGTCTGAGCGCGGCTGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB 757 TTCAACCTCGAGCGCGGCTGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
QY 841 GCGGCTCTCGGCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
DB 817 GCGGCTCGAGATGAGTCCAGCGGCGGCTCTGAGCGGCGGCGGCGGCGGCGGCGG 876
QY 901 TCGAGGCGGCTCAAGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
DB 877 TCGAGTGTCTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
QY 961 CCNTTCTCGACGAGCTGACCGGCGGCTCACCGGCGGCGGCGGCGGCGGCGGCGG 1020
DB 937 CCNTTCTCGACGAGCTGACCGGCGGCTCACCGGCGGCGGCGGCGGCGGCGGCGG 996
QY 1021 CGCAGCTCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
DB 997 CGCAGCTCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
QY 1081 GACGAGGAGGCGGATTTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1057 GACGAGGAGGCGGATTTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
QY 1141 GCGCGGCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
DB 1117 GCGCGGCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
QY 1201 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
DB 1177 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1236
QY 1261 TTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
DB 1237 TTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1296
QY 1321 GCGAGCTCTCACTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
DB 1297 GCGAGCTCTCACTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1356
QY 1381 CCNTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
DB 1357 CCNTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
QY 1441 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
DB 1417 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
QY 1501 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
DB 1477 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536
QY 1561 ACNCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
DB 1537 ACNCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1596
QY 1621 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
DB 1597 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1656
QY 1681 CCNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
DB 1657 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716
QY 1741 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
DB 1717 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1776
QY 1801 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
DB 1777 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1836
```


QY	1	AATGGGCCCCCTCMTGACATCTGATCGCAGACCTCAATCGCAACGGGANTCNCAACGANTG	60
QY	1	AATGGACCCCTCGCTCGACTGATCTGCGACTCTGTTCCGAAACGGCGGTGCATCCGCAAG	60
Db	1	ATGGCAACCCCTCGCTCGACTGATCTGCGACTCTGTTCCGAAACGGCGGTGCATCCGCAAG	60
QY	61	CACGCCGNNCGCNCNNNCGACNCAACGGGCGCCACATCTCAACNTCNCGCCGCGCG	120
Db	61	CAGGCTGT-----CAATGGCGCCTCTGCACCAACCTC-----GCACTTCGAG	100
QY	121	GCTCGCTCTCTCCCGACCAACCCAGNNNGACGACACTGCATCGTGNAGNAGTCTCTCGCG	180
Db	101	GCTCGACACTCGCCACCAACCCAGGTGCACGAGAGTGCACATGTCGAAAGATGCTCGCG	160
QY	181	ACCCCAACCGNNACGACGNNNTTGAACTTCAGACGGGTACACCTTACCTTGGGNAAGTCG	240
Db	161	CGCCGACCGACTCGA--CGCTCGAACTTCGACGGGTACTCGCTAACCTTGGGNAAGTCG	217
QY	241	TGCGCGCCGCGNCGCAAGGAGCGCGCNCNGTCCCGCTCNCAAGACNCGACGAGATCCGGCA	300
Db	218	TCTCGCGCCGCGAGGAAGGGCAGGCGCTGTCCCGCTCAGAGAGAG--CGACGAGATCCGCTCA	276
QY	301	AAGATTCGACAAANAGTTCGAGTTCTCTCCGANNCAAGCTCNAACAAAGNCTTCACGNGTC	360
Db	277	AAGATTGCAAAATCGGTTCGAGTTCTTGGGCTCGGAATCTTCCAGAGACGCTTCACGGGTC	336

QY	361	ACGACGTGGTTGCGGCGCTCGCGCCGACACCGGACTGAGGATATGCNATCTCCGCTCCGAGAG	420
Db	337	ACGACTGGATTGCGGGAATCCGCAAGACCCCGACACCGAGGCGCATCTCGCTCCAGAGG	396
QY	421	GCNCTCTCTGAGCACAGACTCTGCGGTGTCCTCCNACGTGNTGCAMTCTTTCNAGCTTC	480
Db	397	GCCTCTCTGAGCACCTCTCTGCGGTGTTCTCCCTTCGTCGATTCGATCTGTTCCGCTTC	456
QY	481	GGNCGGCGCTTCGAGAAATCGCTTCCGCTGGAAGTGTCCGCGGCGCCATGACATCCGC	540
Db	457	GGCGGCGGTTCGAGAAATCGCTTCCCTCGAAGTGTTCGCGGCGCCATGACAAATCCGC	516
QY	541	GTCAACTGACTACACGCGGCACTCGCGCNGTCCGCTCGTCTGCTCCAGGCGCTACAC	600
Db	517	GTCAAAGATTGACCCCGGCGCACTCGGCTGTCCGCTCGTCTCGAGGCGCTACAC	576
QY	601	AACTTCCTCTAACACACGCGATACCCCATGTCTCCCTTCGCGGCGCACTCTCGCGCTGC	660
Db	577	AACTTCCTCTAACACACGCGATACCCCATGTCTCCCTTCGCGGCGCACTCTCTCGCTGC	636
QY	661	GGCGACTCTCCCGCCTCTGTCATCGCGCGCGCATACCGGTATACCCCGGACNNCAAG	720
Db	637	GGCGACTCTCTCTCTCTCTCTTAATTGACGGGCACTACGCGGTACCCGAGACAGAG	696
QY	721	GTNCACGTNNTNCAACGAGGCGANNAGAGATCATGTNCGCCGCGAGGCGATTCGCGCTC	780
Db	697	GTGCAACGTCTGTCCACGAGGCGAAGAAATCTGTACGCGCGGACGATGAGCGCTC	756
QY	781	TNNGGTCTCGAGCCCGCTCTCTCTCGGCGCCAAAGAGGTCTCGGTCTGTCAACGCGACG	840
Db	757	TTCAACTCTGAGCCCGCTCTCTCTGCGCGCAAGAGAGGTCTCGGTCTGTCAACGCGACG	816
QY	841	GCGCTCTCGGCTCGATGAGCGACCCCTCGCTCTCAAGAGCAACATGCTCTCGCTCTC	900
Db	817	GCGCTCTGAGATGAGATGAGCGACCCCTCGCTCTCAAGAGCAACATGCTCTCGCTCTC	876
QY	901	TGCGAGCGCTACCGGCTCTTACGCGTTCGAGGCGATGTCGCGCCACGCGGCTCTCTCAC	960
Db	877	TGCGAGTGTCTACCGGCTCTGACGCGTTCGAGGCGATGTCGCGCCACGCGGCTCTCTCAC	936
QY	961	CGNTTCCTCCAGACGTCACGCGCGCTCAACCGGACCCCAATGAGAGTCCGCGCAACATC	1020
Db	937	CGCTTCCTTTCACGACGTACGCGCTCAACCGGACCCCAATGAGAGTCCGCGGAAATCTC	996
QY	1021	CGCAGCGTCTCGAGGCGACGNNGTTTGCGGTTCACACAGAGAGAGGTCAAGGTCAAG	1080
Db	997	CGCAGCGTCTCGAGGCGACGCGTTGTGCTTCACCATGAGAGAGAGGTCAAGGTCAAG	1056
QY	1081	GACGACGAGGCGCATTTCTCCGCGAGAGACCGGTACCCGCTCCGACGTCCGCTCAATGCGTCTC	1140
Db	1057	GACGACGAGGCGCATTTCTCCGCGAGAGACCGGTACCCCTTGCGACGTCTCTCAATGCGTCTC	1116
QY	1141	GGCCCGCTCTCGACGCAATGATTTCAAGCCCAACGCGTCTCTCGCTCGAGGCGAGCAG	1200
Db	1117	GGCCCGCTCTCGACGCAATGATTTCAAGCCCAACGCGTCTCTCAACATGAGGCGCGCAG	1176
QY	1201	TGCGAGACCGCAAAACCCGCTCATCTGACGTTCGAGAACAGAGAACCCACACGCGCGCAAC	1260
Db	1177	TGCGAGACCGCAAAACCCCTTCATCTGACGTTCGAGAACAGACTTTCACACACGCGCGCAAT	1236
QY	1261	TTCGAGAGGAGGCGGTGTCGAGNAAACGATGAGAGAAACATCGGCTCGGCTCGCCCTGATC	1320
Db	1237	TTCGAGAGGTCGCGTGTGCGCAACCACTGAGAGAACTCGGCTCGGCTCGCCCAATC	1296
QY	1321	GGCAAGCTCAACTTCACGCACTTACCGGATGCTCAACGCGGAGTGAACCGCGGCTTN	1380
Db	1297	GGCAAGCTCAACTTCACGCACTTACCGGATGCTCAACGCGGAGTGAACCGGCGCTC	1356
QY	1381	CGCTCTGCTCTCGTGGCGAGACCGTCTCTCTTACATCTGCAAGGCTCGACATTT	1440
Db	1357	CGCTCTGCTCTCGTGGCGAGACCGCTCTCTCTTACATCTGCAAGGCTCGACATCT	1416

QY 1441 GCCGCGCNGCAGTACCTTGGAGCTCGAGCACTTGGCAAACCGGTNAGACCCACGTC 1500
DB 1417 GCCGCTGGCGCGTACACCTCGAGGTGGGACACTGGCCAAACCTGTGACGACGATGTC 1476
QY 1501 CAGCCGCGCNGAGATGGGCAACGAGCGGTCAACTGCTCGCCTGATCTCGGCGCGCCG 1560
DB 1477 CAGCCGCGCTGAGATGGCGAAGCGGCTCAACTGCTGCGCTCATCTCGGCTGTGCG 1536
QY 1561 ACNCGCGAGGCGCAACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1537 ACAGCCGAGTCCACACAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGGCGCTGACCTCGCGCGATGAGTTGAGTTCAAGAGCACTTCAACCCGNTNTC 1680
DB 1597 CAGGCGATGACTTGGCGCGCGAGCGGTTGAGTTCAAGAGCACTTGGCCAGCATC 1656
QY 1681 NCNNGCTCTCTAGAGAGACTTGGCNGCGCTCGAGCGCNGNNNNNNNNNNNACGAA 1740
DB 1657 GTCTGCTCATGACAGCACTTTGGCTCGGCTGACGAGCTGACCTGGCGGAGAG 1716
QY 1741 CTCGNGGACAGGTCAACAAGNCGCTCNAACAAGCNGCTGAGCAGACCACTCGTACGAC 1800
DB 1717 CTGCTGAGAGAGGTGAAGAGAGCGTGGCAAGCGCTGAGCAGACCACTCGTACGAC 1776
QY 1801 CTGAGCGCGCTGGGCAAGCGCTTCTGTTGCGGACCGGACCGCTGCTGAGNNNNN 1860
DB 1777 CTGCTCGCGCTGGGCAAGCGCTTCTGTTGCGGCGCGGACCGCTGAGGTTCTC 1836
QY 1861 NNCTCTCGCNGNNNGCANNAGGTCTGCTCGCNGCGCTCAAGCGCTGAAAGTGGCT 1920
DB 1837 TCCTGACGTC-----CTCTCGCTGCGCGCGCTCAAGCGCTGAAAGTGGCG 1885
QY 1921 CCGCGGAGAAAGCCATCTGCTCAAGCGCANGTCCGGGACNCTTCTGAGCGGCGCT 1980
DB 1886 CCGCGGAGTGGGCAATCTGCTCAAGCGCANGTCCGGGACNCTTCTGAGTCCGCGCT 1945
QY 1981 CGTGTGTGTGCGCGGCTCNGGTACTCTGCGCGGCGACGCGGCTCTGTACTGTGCTG 2040
DB 1946 CGACTGTGTGCGCGGCTCTGTACTCTGCGCGGCACTGAGTCTCTACGCTTTCG 2005
QY 2041 TCGCGAGAGAGCTCGGCGTCAAGGCGCGCGGCGAGAGTCTTCTCTCGCAAGGAGAG 2100
DB 2006 TCGCGAGAGAGCTTGGGCTCAAGGCGCGCGGAGAGAGTCTTCTCTCGCAAGGAGAG 2065
QY 2101 TGACGATCGGCAACAGTCTCCGCACTTACGAGGCACTCAAGNNCGGCGCATCAAC 2160
DB 2066 TGACGATCGGCTGAAAGCTCTCAAGATCTACGAGGCACTCAAGTGGGCGAGATCA 2125
QY 2161 ACCGTCTGTCAAGATGCTGAGTAG 2186
DB 2126 ACGTCTCTCAAGATGCTGAGTAG 2151

Search completed: September 12, 2004, 12:22:21
Job time : 1619.68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:01:03 ; Search time 23.5975 Seconds
(without alignments)
1598.322 Million cell updates/sec

Title: US-09-939-408a-21
Perfect score: 3365
Sequence: 1 MAPSDSTATSXANGXNGX.....RIYAIKGRINXVYVLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6a_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6b_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3223	95.8	726	4 US-09-624-693A-21	Sequence 21, App1
2	3021	89.8	716	4 US-09-624-693A-19	Sequence 19, App1
3	3021	89.8	716	4 US-09-627-216A-8	Sequence 8, App1
4	3021	89.8	716	4 US-09-765-873A-8	Sequence 8, App1
5	3019	89.7	716	4 US-09-627-216A-10	Sequence 10, App1
6	3019	89.7	716	4 US-09-765-873A-10	Sequence 10, App1
7	3014	89.6	716	4 US-09-765-873A-38	Sequence 38, App1
8	3012	89.5	716	4 US-09-765-873A-35	Sequence 35, App1
9	3010	89.5	716	4 US-09-765-873A-36	Sequence 36, App1
10	3010	89.5	716	4 US-09-765-873A-37	Sequence 37, App1
11	3006	89.3	716	4 US-09-765-873A-34	Sequence 34, App1
12	2978	88.5	716	4 US-09-765-873A-33	Sequence 33, App1
13	2971	88.3	686	4 US-09-765-873A-32	Sequence 32, App1
14	2922.5	86.8	713	4 US-09-624-693A-17	Sequence 17, App1
15	2770	82.3	720	4 US-09-624-693A-13	Sequence 13, App1
16	1294.5	38.5	740	4 US-09-624-693A-15	Sequence 15, App1
17	670	19.9	424	4 US-09-615-192A-128	Sequence 128, App
18	495.5	14.7	515	4 US-09-328-352-5090	Sequence 5090, App
19	490	14.6	508	4 US-09-489-039A-13575	Sequence 13575, A
20	480.5	14.3	515	4 US-09-252-991A-13575	Sequence 13575, A
21	474.5	14.1	525	4 US-09-540-236-2855	Sequence 2855, App
22	459	13.6	412	4 US-09-615-192A-130	Sequence 130, App
23	365	10.8	520	4 US-09-252-991A-13049	Sequence 13049, A
24	223	6.9	226	4 US-09-615-192A-327	Sequence 327, App
25	223.5	6.6	97	4 US-09-615-192A-329	Sequence 329, App
26	115	3.4	1024	4 US-09-562-737-50	Sequence 50, App
27	111	3.3	1024	4 US-09-562-737-41	Sequence 41, App1

28	108	3.2	1024	4 US-09-562-737-44	Sequence 44, App1
29	106	3.2	505	4 US-09-252-991A-23925	Sequence 23925, A
30	104	3.1	744	4 US-08-764-100-25	Sequence 25, App1
31	102.5	3.0	1011	4 US-09-252-991A-22505	Sequence 22505, A
32	100.5	3.0	555	4 US-09-543-681A-7878	Sequence 7878, App
33	100	3.0	1589	4 US-09-543-681A-4958	Sequence 4958, App
34	99.5	3.0	1024	4 US-09-562-737-45	Sequence 45, App1
35	96.5	2.9	404	4 US-09-328-352-6480	Sequence 6480, App
36	96	2.9	875	4 US-09-328-352-4884	Sequence 4884, App
37	95.5	2.8	836	4 US-09-252-991A-23513	Sequence 23513, A
38	94.5	2.8	607	4 US-08-969-683A-59	Sequence 59, App1
39	94	2.8	555	4 US-08-440-317A-6	Sequence 6, App1
40	94	2.8	555	2 US-08-687-852-6	Sequence 6, App1
41	94	2.8	764	4 US-09-252-991A-31816	Sequence 31816, A
42	93.5	2.8	1024	4 US-09-562-737-47	Sequence 47, App1
43	93.5	2.8	1024	4 US-09-562-737-49	Sequence 49, App1
44	93.5	2.8	1495	4 US-08-522-726B-1	Sequence 1, App1
45	93.5	2.8	1495	4 US-09-337-384-1	Sequence 1, App1

ALIGNMENTS

```
RESULT 1
US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
; OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
US-09-624-693A-21

Query Match      95.8%; Score 3223; DB 4; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MAPSDSTATSXANGXNGXAAAXXXXXXXXAGSXLPETXXXTOLDIVEXXLD 60
        1 MAPSDSTATSXANGXNGXAAAXXXXXXXXAGSXLPETXXXTOLDIVEXXLD 60
DB      1 MAPSDSTATSXANGXNGXAAAXXXXXXXXAGSXLPETXXXTOLDIVEXXLD 60
QY      61 PXTDXXHEIDGYSLTIGDVGAAKGRXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
        61 PXTDXXHEIDGYSLTIGDVGAAKGRXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
DB      61 PXTDXXHEIDGYSLTIGDVGAAKGRXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
QY      121 TFGGSAATREDIASLOKALLEHQLCGVLTSDXSPXKLGLENSLPLEVVGAMITRV 180
        121 TFGGSAATREDIASLOKALLEHQLCGVLTSDXSPXKLGLENSLPLEVVGAMITRV 180
DB      121 TFGGSAATREDIASLOKALLEHQLCGVLTSDXSPXKLGLENSLPLEVVGAMITRV 180
QY      181 NSLTGSHAVRLVVEALTNFLNGITPIVPLRGITIASGDLSPSYIAAITGHPGSXV 240
        181 NSLTGSHAVRLVVEALTNFLNGITPIVPLRGITIASGDLSPSYIAAITGHPGSXV 240
DB      181 NSLTGSHAVRLVVEALTNFLNGITPIVPLRGITIASGDLSPSYIAAITGHPGSXV 240
QY      241 HXHEGXKIXKARBAIALFGLPEPVYLPKKGGLGVNGTAVASAMATLALDAHMSLLS 300
        241 HXHEGXKIXKARBAIALFGLPEPVYLPKKGGLGVNGTAVASAMATLALDAHMSLLS 300
```

Db	24	HYHGEKEMKAREALFGLPEVYLGPEKGLGVNQTAVSASVATLADHMTLSLS	300
Qy	301	QALTALTVEAMVGHASFFHPFLHDVTRHPFQIIEVARNIRTLLEGSPYAHNEEEVYKD	360
Db	301	QALTALTVEAMVGHASFFHPFLHDVTRHPFQIIEVARNIRTLLEGSPYAHNEEEVYKD	360
Qy	361	DEGILKQDRPLRTSPQWGLPVSQWIMHAHVLSEAGOSTTNDPLIDVENKXTHHGNE	420
Db	361	DEGILKQDRPLRTSPQWGLPVSQWIMHAHVLSEAGOSTTNDPLIDVENKXTHHGNE	420
Qy	421	QASAVXNTEKTRLTALALIGKLNFTQLTETMLNAGNRCGLPSCLAEDPDSLSYHCKGLDIA	480
Db	421	QASAVXNTEKTRLTALALIGKLNFTQLTETMLNAGNRCGLPSCLAEDPDSLSYHCKGLDIA	480
Qy	481	AAAYTSELGLNAPVTVTHQPAEMNGQANSLALSARTAEADVLSLLATLTVLVQ	540
Db	481	AAAYTSELGLNAPVTVTHQPAEMNGQANSLALSARTAEADVLSLLATLTVLVQ	540
Qy	541	AVDLRAMEFEFKQEPFYXXXLLXOHFGXXTXXXXXELXXKYKXLKRLAEQTSNYDL	600
Db	541	AVDLRAMEFEFKQEPFYXXXLLXOHFGXXTXXXXXELXXKYKXLKRLAEQTSNYDL	600
Qy	601	EPFMHDAFSXATGTVEEXLSYXXXXXVSLAAVANAKVAXAEKALSLTRVYXXFWXAPSS	660
Db	601	EPFMHDAFSXATGTVEEXLSYXXXXXVSLAAVANAKVAXAEKALSLTRVYXXFWXAPSS	660
Qy	661	SSPALXYLSPTTGVLYSFFREELGVKARGDVFLGKQSVTIGSNVSRVYEAIXGRINXY	720
Db	661	SSPALXYLSPTTGVLYSFFREELGVKARGDVFLGKQSVTIGSNVSRVYEAIXGRINXY	720
Qy	721	LVKMLA 726	
Db	721	LVKMLA 726	

RESULT 2
US-09-624-693A-19
Sequence 19, Application US/09624693A
Patent No. 6355468
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Kocotetra, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NCC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 716
TYPE: PRT
ORGANISM: Rhodococcus toruloides
US-09-624-693A-19

Query Match	89.8%	Score 3021	DB 4	Length 716
Best Local Similarity	84.0%	Pred. No. 0		
Matches 610	Conservative 25	Mismatches 81	Indels 10	Gaps 3

QY	1	MAPSLSDIATSYANGXNGXHAAXXAXXXXXXXXAXSGLPTTXXTQGLDIVEXXLAD	60
Db	1	MAPSLSLSISHSPANGVASAKQAVNGAS-----TNLVAAGSHLPTTQVQVDIVEXMLAA	54
QY	61	PXIDDXAELEDDGSLTSLGDTVGAARGRGRYRXDSDEFRTXITDKSVETLRQLXNSRYGTT	120
Db	55	P-PDSTLELDGSLNMGDVSAARGRPRYXDSDEFRTSKIDSVETLRQLSMSYGYGT	113
QY	121	TGFGGSADPRTEDATSLQKALLLEHQLCVLTPTSDXSYFLRGLENSLPLEVVGAMTTIRY	180
Db	114	TGFGGSADPRTEDATSLQKALLLEHQLCVLTPTSDPRLRGLENSLPLEVVGAMTTIRY	173
QY	181	NSITRGHSAVRLVLEALTNPLNHSITPIVPLRGITISAGDSBLSYIAAALTGHPSKY	240

Db	174	NSLTRSHASVRLVVLDEALTNFLNHGITDIVEFLRGTSASGDLSPLSYIAAASGHPDSKV	2333
Qy	241	HYXHEXEXIMARAREPAILFGLEPVLVGRKXGLVNGTANSASMATLALHDAMLSLS	3000
Db	234	HVHEKEKELILAREEMALFNLEPVLGPKGLVNGTAASAMATLALHDAMLSLS	2933
Qy	301	QALTLTVAMYGHAGSFHPLHDVTRPHPTQIEVARNIRPLEGSXFAVHHEEVKVD	3600
Db	294	QSLTANTVAMYGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSXFAVHHEEVKVD	3533
Qy	361	DEGLIFRDRYPLRTSPQMLGPIVSDMIHAHAVISLEAGOSTTDPNPLIDVENKKTTHGNGF	4200
Db	354	DEGLITRQDRYPLRTSPQMLGPIVSDLIHAHAVISLEAGOSTTDPNPLIDVENKKTSHHGNGF	4133
Qy	421	QASAVYNTKTRLALALIGKLNPFQLEMNAGMNGRLPSCLAAPDPSLSYHCKGLDIA	4800
Db	414	QPAAVYNTKTRLGLAQIGKLNPFQLEMNAGMNGRLPSCLAAPDPSLSYHCKGLDIA	4733
Qy	481	AAAYTSELGLANPYTHVQPAEMGNQAVNSLALISARTEANDVLSLLATHLXCVLQ	5400
Db	474	AAAYTSELGLANPYTHVQPAEMVANOAVNSLALISARTTESNDVLSLLATHLXCVLQ	5333
Qy	541	AUDLRAMEEFPKQKQXPXXXLIXHPFXKXTXXXXXXELXKTKXKXLRLQOTNSYDL	6000
Db	534	AUDLRABEEFPKQKQGPALVSLIDHFSASMTGSLDELVEKNKTLARLEQOTNSYDL	5933
Qy	601	EPRWHAFAFXTAGTVEEXLSXXXXXVSLAAMNMTAKAAEKAISLTRYRXKFWXAPS	6600
Db	594	VPRWHAFAFPAAGTVEVLS---TSLSLAAMNMTAKAAEASISLRQVRETFWMAST	6500
Qy	661	SSPALXYLSPRTVLVYSVREBELGVKARGDVFLGKQVTTGSVNSFIYALISGRINXY	7200
Db	651	SSPALXYLSPRTQIILYAFVREBELGVKARGDVFLLGKQVTTGSVNSKIYEAISGRINNY	7100
Qy	721	LVYQMLA 726	
Db	711	LVYQMLA 716	

```

RESULT 3
US-09-627-216A-8
/ Sequence 8: Application US/09627216A
/ Patent No. 6368837
/ GENERAL INFORMATION:
/ APPLICANT: Sartislati, Sima F
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ APPLICANT: Vannelli, Todd
/ APPLICANT: Garenby, Anthony
/ TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
/ FILE REFERENCE: BC1009 US NA
/ CURRENT APPLICATION NUMBER: US/09/627.216A
/ PRIOR FILING DATE: 2000-07-27
/ PRIORITY FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 8
/ LENGTH: 716
/ TYPE: PRP
/ ORGANISM: Rhodotorula glutinis
US-09-627-216A-8

Query Match      89.8%; Score 3021; DB 4; Length 716;
Best Local Similarity 84.0%; Fred. No. 0;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3

1 MAPSLDIATSYXANGXNGXHAAXASXXXXXXXXXXAGSXPPTXXTQLDIVEXXLAD 60
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
1 MAPSLDISHSANVNAKAVNAGS-----TNLAVAOSHLPPTQTGVQVDIVEKGLAA 54
Db
Yy          61 PXTDDXHELDGYSITLGGVAVAAKGRXRVXDSDERFKXIDSVFLEKQLXNSRYGT 120

```

Db 55 P-TDSTLELDGSLNIGDVASARKGRPVXKDSDEIRSKIDKVEFRLSOLNSVYGV 113
QY 121 TGFSGSADTRTEDAISLQKALLEHQLCVLPFSXDSFXLGRLENSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCVLPFSXDSFXLGRLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGHSARLVVLEALTNFNLHGITPIVPLRGITTSASGDISPLSYIAAITHGPD SKV 240
Db 174 NSLTRGHSARLVVLEALTNFNLHGITPIVPLRGITTSASGDISPLSYIAAISHGPD SKV 233
QY 244 HYHHEXEXKIMKAREALFGLPEVVLGPKGGLGVNGTAVASAMATLALHDAMLSTLS 300
Db 234 HYHHEXEXKILYAREMALLFNLEPVVLGPKGGLGVNGTAVASAMATLALHDAMLSTLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSXFVYHHEEVKVD 360
Db 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSREFAVHHEEVKVD 353
QY 361 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGNE 420
Db 354 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGNE 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRRLAQIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLIATLHYCYLQ 540
Db 474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAESNDVLSLLIATLHYCYLQ 533
QY 541 AYDLRAMEPEFKQKQFXXXXLXOHFGXXXXXELXXXXKXKXLEOTNSYDL 600
Db 534 AIDLRARIEEFKQKQFPAIVSLIDQFSGAMTGSNLRDELVEKNKTLAKLEOTNSYDL 593
QY 601 EPRMHDAPFSAATGTVVEXLSXXXXXVSLAAVNAMKVAAXEKAISLTRVXXFXMXPSS 660
Db 594 VPRMHDAPFSAATGTVVEXLS--TSLSLAAVNAMKVAAXEKAISLTRQVETFWMSAST 650
QY 661 SSPALXYLSPTRTVLYSFVREBELGYKARRGDVFLGKQVTTGNSVSKRYEAIKSGRINXV 720
Db 651 SSPALXYLSPTRTQILYAFVREBELGYKARRGDVFLGKQVTTGNSVSKRYEAIKSGRINXV 710
QY 721 LVKMLA 726
Db 711 LVKMLA 716

RESULT 4

US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIORITY FILING DATE: 2000-01-19
; PRIORITY APPLICATION NUMBER: US 09/627, 216
; PRIORITY FILING DATE: 2000-07-27
; PRIORITY APPLICATION NUMBER: US 60/147, 719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula glutinis
US-09-765-873A-8

Query Match 89.8%; Score 3021; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDSTATSXANGXNGXHAAXXXXXXXXAXAGSLPTTXXXTQDLIVEXKLIAD 60
Db 1 MAPSLDSTATSXANGXNGXHAAXXXXXXXXAXAGSLPTTXXXTQDLIVEXKLIAD 54
QY 61 PYTDXXELDLGSLNIGDVGAARKGRPVXKDSDEIRSKIDKVEFRLSOLNSVYGV 120
Db 55 P-TDSTLELDGSLNIGDVASARKGRPVXKDSDEIRSKIDKVEFRLSOLNSVYGV 113
QY 121 TGFSGSADTRTEDAISLQKALLEHQLCVLPFSXDSFXLGRLENSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCVLPFSXDSFXLGRLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGHSARLVVLEALTNFNLHGITPIVPLRGITTSASGDISPLSYIAAITHGPD SKV 240
Db 174 NSLTRGHSARLVVLEALTNFNLHGITPIVPLRGITTSASGDISPLSYIAAISHGPD SKV 233
QY 244 HYHHEXEXKIMKAREALFGLPEVVLGPKGGLGVNGTAVASAMATLALHDAMLSTLS 300
Db 234 HYHHEXEXKILYAREMALLFNLEPVVLGPKGGLGVNGTAVASAMATLALHDAMLSTLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSXFVYHHEEVKVD 360
Db 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSREFAVHHEEVKVD 353
QY 361 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGNE 420
Db 354 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGNE 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRRLAQIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLIATLHYCYLQ 540
Db 474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAESNDVLSLLIATLHYCYLQ 533
QY 541 AYDLRAMEPEFKQKQFXXXXLXOHFGXXXXXELXXXXKXKXLEOTNSYDL 600
Db 534 AIDLRARIEEFKQKQFPAIVSLIDQFSGAMTGSNLRDELVEKNKTLAKLEOTNSYDL 593
QY 601 EPRMHDAPFSAATGTVVEXLSXXXXXVSLAAVNAMKVAAXEKAISLTRVXXFXMXPSS 660
Db 594 VPRMHDAPFSAATGTVVEXLS--TSLSLAAVNAMKVAAXEKAISLTRQVETFWMSAST 650
QY 661 SSPALXYLSPTRTVLYSFVREBELGYKARRGDVFLGKQVTTGNSVSKRYEAIKSGRINXV 720
Db 651 SSPALXYLSPTRTQILYAFVREBELGYKARRGDVFLGKQVTTGNSVSKRYEAIKSGRINXV 710
QY 721 LVKMLA 726
Db 711 LVKMLA 716

RESULT 5

US-09-627-216A-10
; Sequence 10, Application US/09627216A
; Patent No. 636837
; GENERAL INFORMATION:
; APPLICANT: Satiaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BCI009 US NA
; CURRENT APPLICATION NUMBER: US/09/627, 216A
; PRIORITY FILING DATE: 2000-07-27
; PRIORITY APPLICATION NUMBER: 60/147, 719
; PRIORITY FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10

LENGTH: 716
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: mutant from
US-09-627-216a-10

Query Match 89.7%; Score 3019; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

```
QY 1 MAPSLDSTATSANGKXGHAAXXAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXAD 60
D 1 MAPSLDSTATSANGKXGHAAXXAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXAD 54
QY 61 PXTDDXXELDGYSLTIGDVVGAARKGRVYXDSDEIRKIDKSVYEFRLXQLXNSVYGV 120
D 55 P-TDSTELDGYSLNIGDVVGAARKGRVYXDSDEIRKIDKSVYEFRLXQLXNSVYGV 113
QY 121 TGFSGADRTEDALSLQKALLEHQCGVLPSTXDSFYLGRGLNSLPLEVVGAMTIRV 180
D 114 TGFSGADRTEDALSLQKALLEHQCGVLPSTXDSFYLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISASGDLSPSTYAAATGHPDSKV 240
D 174 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISASGDLSPSTYAAATGHPDSKV 233
QY 241 HVXHEGKEKIMXAREALFGLPEVVLGPKGGLVNGTAVSASMATLALDHAMLSLS 300
D 234 HVXHEGKEKILYAREAMALFNLPEVVLGPKGGLVNGTAVSASMATLALDHAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHFLHDVTRPHPTQIEVANIRITLLEGSXFAVHEEYVKD 360
D 294 QSLTAMTYEAMVGHAGSFHFLHDVTRPHPTQIEVANIRITLLEGSXFAVHEEYVKD 353
QY 361 DGLIRQDRYPLRTSPQMLGVLSDMIHAHVALISLEAGOSTTDNPLIDVENKXTHHGNF 420
D 354 DGLIRQDRYPLRTSPQMLGVLSDMIHAHVALISLEAGOSTTDNPLIDVENKXTHHGNF 413
QY 421 QASAVNMTKTRIALALIGKLNFTQLTMLNAGNRGLPSCLAEDPSLSYHCKGLDIA 480
D 414 QAAAVNMTKTRIALALIGKLNFTQLTMLNAGNRGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTTHVOPAEANOAVNSLALISARTAEANDVLSLLATHLYCVLQ 540
D 474 AAAYTSELGHLANPVTTHVOPAEANOAVNSLALISARTTESNDVLSLLATHLYCVLQ 533
QY 541 AVDLRAHEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXXKXLEKLEQNSYDL 600
D 534 AIDLRATEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXXKXLEKLEQNSYDL 593
QY 601 EPRMHDASXAGTVEEVLSSXXXXXSLAAVNAKVAEKAISLITXVXXFXKAPSS 660
D 594 VPRMHDASFAAGTVEEVLSS--TSLSLAAVNAKVAEKAISLITRQVETFEFSAAST 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGKQEVLTIGSNVSRITYEAIKSGRINXV 720
D 651 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGKQEVLTIGSNVSRITYEAIKSGRINXV 710
QY 721 LVKMLA 726
D 711 LKMLA 716
```

RESULT 6
US-09-765-873a-10
Sequence 10, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOREDUCTION OF PABA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US CIP

CURRENT APPLICATION NUMBER: US/09/765, 873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627, 216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147, 719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 716
TYPE: PRT
ORGANISM: mutant from Rhodocorula glutinis
US-09-765-873a-10

Query Match 89.7%; Score 3019; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

```
QY 1 MAPSLDSTATSANGKXGHAAXXAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXAD 60
D 1 MAPSLDSTATSANGKXGHAAXXAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXAD 54
QY 61 PXTDDXXELDGYSLTIGDVVGAARKGRVYXDSDEIRKIDKSVYEFRLXQLXNSVYGV 120
D 55 P-TDSTELDGYSLNIGDVVGAARKGRVYXDSDEIRKIDKSVYEFRLXQLXNSVYGV 113
QY 121 TGFSGADRTEDALSLQKALLEHQCGVLPSTXDSFYLGRGLNSLPLEVVGAMTIRV 180
D 114 TGFSGADRTEDALSLQKALLEHQCGVLPSTXDSFYLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISASGDLSPSTYAAATGHPDSKV 240
D 174 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISASGDLSPSTYAAATGHPDSKV 233
QY 241 HVXHEGKEKIMXAREALFGLPEVVLGPKGGLVNGTAVSASMATLALDHAMLSLS 300
D 234 HVXHEGKEKILYAREAMALFNLPEVVLGPKGGLVNGTAVSASMATLALDHAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHFLHDVTRPHPTQIEVANIRITLLEGSXFAVHEEYVKD 360
D 294 QSLTAMTYEAMVGHAGSFHFLHDVTRPHPTQIEVANIRITLLEGSXFAVHEEYVKD 353
QY 361 DGLIRQDRYPLRTSPQMLGVLSDMIHAHVALISLEAGOSTTDNPLIDVENKXTHHGNF 420
D 354 DGLIRQDRYPLRTSPQMLGVLSDMIHAHVALISLEAGOSTTDNPLIDVENKXTHHGNF 413
QY 421 QASAVNMTKTRIALALIGKLNFTQLTMLNAGNRGLPSCLAEDPSLSYHCKGLDIA 480
D 414 QAAAVNMTKTRIALALIGKLNFTQLTMLNAGNRGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTTHVOPAEANOAVNSLALISARTAEANDVLSLLATHLYCVLQ 540
D 474 AAAYTSELGHLANPVTTHVOPAEANOAVNSLALISARTTESNDVLSLLATHLYCVLQ 533
QY 541 AVDLRAHEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXXKXLEKLEQNSYDL 600
D 534 AIDLRATEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXXKXLEKLEQNSYDL 593
QY 601 EPRMHDASXAGTVEEVLSSXXXXXSLAAVNAKVAEKAISLITXVXXFXKAPSS 660
D 594 VPRMHDASFAAGTVEEVLSS--TSLSLAAVNAKVAEKAISLITRQVETFEFSAAST 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGKQEVLTIGSNVSRITYEAIKSGRINXV 720
D 651 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGKQEVLTIGSNVSRITYEAIKSGRINXV 710
QY 721 LVKMLA 726
D 711 LKMLA 716
```

RESULT 7
US-09-765-873a-38

```

; Sequence 38, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
; LENGTH: 716
; TYPE: PRP
; ORGANISM: mutant from Rhodocorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: X = Gly, Ala, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X = Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match      89.6%; Score 3014; DB 4; Length 716;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

QY      1 MAPSLDSTISXANGXNGKHAAXASXXXXXXXAGXSLTXXQTQDIYEXXLD 60
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MAPSLDSTISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTDYVERKMLAA 54
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      61 PXTDXXLDDGSLTLGDIVGAARKGRVYRXDSDERXKIDKSVFRLXQLXNSVYGT 120
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      55 P-TDSTLELDGSLNLGDVVSAAKGRVYRXDSDERISKIDKSVFRLSGLSWSYGT 113
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      121 TGFSGADTRTEDAISLOKALLEHQLCGVLPTSXDSFXLGRLENSLPLEVVRGAMTIRV 180
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      114 TGFSGADTRTEDAISLOKALLEHQLCGVLPTSXDSFXLGRLENSLPLEVVRGAMTIRV 173
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      181 NSLTRGHSNRLVVEALTNFLNGITIPVLRGTISASGDLSPSTYIAAATGHPPSKV 240
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      174 NSLTRGHSNRLVVEALTNFLNGITIPVLRGTISASGDLSPSTYIAAATGHPPSKV 233
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      241 HXHEGXEKIMKAREBAIALFGLPEVYLGPKEGLGVNGTAVASAMATLALHDAHMLSLLS 300
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      234 HXHEGXEKILYAREBAMLFNLEPVYLGPKGLGVNGTAVASAMATLALHDAHMLSLLS 293
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRLLGSAFVHHEEVYKD 360
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSAFVHHEEVYKD 353
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      361 DSGILRODRYPRTSPQWGLPVSDMIHAHVLLEAGOSTTNDPLIDVENKXTHHGNF 420
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      354 DSGILRODRYPRTSPQWGLPVSDLIHAHVLLEAGOSTTNDPLIDVENKXTHHGNF 413
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      421 QASAVXNTEKTRIALALIGKLNFTQLTENLNAQNRGLPSCLAEDPSLSYHCKGLDIA 480
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      414 QAAAVANTMEKTRILGAIQIKLNFQLTENLNAQNRGLPSCLAEDPSLSYHCKGLDIA 473
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      481 AAAYTSELGHLNPTVTHYQPAKNGQAVNSIALISARTTAENDVLSLLATLHYCLQ 540
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      474 AAAYTSELGHLNPTVTHYQPAKMANQAKNSIALISARTTESNDVLSLLATLHYCLQ 533
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      541 AAVLRAMEEFKQFPFXXXXLLXQHFQXXXXTXXXXXELXXKXKXKLEQTNSTYL 600
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      534 AILRLAXEFPFKQFPBALVSLIDQHFSGAMGSLRDELVEKNKTLAKRLEQTNSTYL 593
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      601 EPRMHDASFAXGTAVEXLSSXXXXXSLAAVANAKVAAAEKALSLTRVXXKFWAASS 660
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      594 VPRMHDASFAXGTAVEXLSS---TSLSLAAVANAKVAAAEASISLTRYVERTFWGAAST 650
QY      661 SSPALXYLSPRTLYSPVRELEGVKARQVFLGKQEVITGSVSRITYPAISGRINXY 720
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      651 SSPALXYLSPRTQILYAFVRELEGVKARRQDVFLGKQEVITGSVSRITYPAISGRINNV 710
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      721 LVKMLA 726
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      711 LTKMLA 716

RESULT 8
US-09-765-873A-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
; TYPE: PRP
; ORGANISM: mutant from Rhodocorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X = Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X = Thr, Ala, Ser, Pro, Gly
US-09-765-873A-35

Query Match      89.5%; Score 3012; DB 4; Length 716;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

QY      1 MAPSLDSTISXANGXNGKHAAXASXXXXXXXAGXSLPTXXQTQDIYEXXLD 60
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MAPSLDSTISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTDYVERKMLAA 54
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      61 PXTDXXLDDGSLTLGDIVGAARKGRVYRXDSDERXKIDKSVFRLXQLXNSVYGT 120
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      55 P-TDSTLELDGSLNLGDVVSAAKGRVYRXDSDERISKIDKSVFRLSGLSWSYGT 113
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      121 TGFSGADTRTEDAISLOKALLEHQLCGVLPTSXDSFXLGRLENSLPLEVVRGAMTIRV 180
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      114 TGFSGADTRTEDAISLOKALLEHQLCGVLPTSXDSFXLGRLENSLPLEVVRGAMTIRV 173
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      181 NSLTRGHSNRLVVEALTNFLNGITIPVLRGTISASGDLSPSTYIAAATGHPPSKV 240
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      174 NSLTRGHSNRLVVEALTNFLNGITIPVLRGTISASGDLSPSTYIAAATGHPPSKV 233
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      241 HXHEGXEKIMKAREBAIALFGLPEVYLGPKEGLGVNGTAVASAMATLALHDAHMLSLLS 300
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      234 HXHEGXEKILYAREBAMLFNLEPVYLGPKGLGVNGTAVASAMATLALHDAHMLSLLS 293
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRLLGSAFVHHEEVYKD 360
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSAFVHHEEVYKD 353
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      361 DSGILRODRYPRTSPQWGLPVSDMIHAHVLLEAGOSTTNDPLIDVENKXTHHGNF 420
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      354 DSGILRODRYPRTSPQWGLPVSDLIHAHVLLEAGOSTTNDPLIDVENKXTHHGNF 413
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      421 QASAVXNTEKTRIALALIGKLNFTQLTENLNAQNRGLPSCLAEDPSLSYHCKGLDIA 480
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      414 QAAAVANTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCLAEDPSLSHCKGLDIA 473
Qy      481 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANVLSLLATHLYCVLQ 540
Db      474 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANVLSLLATHLYCVLQ 533
Qy      541 AVDIRAMEFEFFKQFPXXXXLXQHFQXXXXTXXXXXELXXXKXKXLXKRLQCTNSYDL 600
Db      534 AIDLRAXEFEFFKQFGPGLVSLIDQHFGSAMTGSNLRDELVEKVKTKLAKLEQCTNSYDL 593
Qy      601 EPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAKVAAEKALSTEXVXXFMPXAPSS 660
Db      594 VPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAKVAAEKALSTEXVXXFMPXAPSS 650
Qy      661 SSPALXYLSPRTVLVSFVREELGVKARRGDVFQKQEVTTGSNVSRIYEAIKSGRLNXY 720
Db      651 SSPALXYLSPRTVLVSFVREELGVKARRGDVFQKQEVTTGSNVSRIYEAIKSGRLNXY 710
Qy      721 LVKMLA 726
Db      711 LKXMLA 716

RESULT 9
US-09-765-873A-36
; Sequence 36, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIORITY FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 2000-07-27
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-36

Query Match      89.5%; Score 3010; DB 4; Length 716;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 608; Conservative 24; Mismatches 84; Indels 10; Gaps 3;

```

```

Db      174 NSLTRGHSAVRLVLEALTNFLNHGITPIVLRGTSISAGDLSPLSYAAALIGHPSKV 233
Qy      241 HYHGXGEXKIXXAAEALFGLEPVVAGKXGLGVNGTAVASAMATLALDHAMLSLS 300
Db      234 HXVHGKEXKILYARAAALFPLEPVVAGKXGLGVNGTAVASAMATLALDHAMLSLS 293
Qy      301 QALTAITYEAMVGHAGSFHPLADYTRPHPTQIEVAAINRTLLGSGFAVHHEEVKXD 360
Db      294 QSLTAMTEAMVGHAGSFHPLADYTRPHPTQIEVAAINRTLLGSGFAVHHEEVKXD 353
Qy      361 DSGILRORVPLRPSQWLGPLVSDMLHAHNVLSLEGGSTTNPPLIDVENKXTHHGNF 420
Db      354 DSGILRORVPLRPSQWLGPLVSDMLHAHNVLSLEGGSTTNPPLIDVENKXTHHGNF 413
Qy      421 QASAVANTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCLAEDPSLSHCKGLDIA 480
Db      414 QAAAVANTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCLAEDPSLSHCKGLDIA 473
Qy      481 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANVLSLLATHLYCVLQ 540
Db      474 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANVLSLLATHLYCVLQ 533
Qy      541 AVDIRAMEFEFFKQFPXXXXLXQHFQXXXXTXXXXXELXXXKXKXLXKRLQCTNSYDL 600
Db      534 AIDLRAXEFEFFKQFGPGLVSLIDQHFGSAMTGSNLRDELVEKVKTKLAKLEQCTNSYDL 593
Qy      601 EPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAKVAAEKALSTEXVXXFMPXAPSS 660
Db      594 VPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAKVAAEKALSTEXVXXFMPXAPSS 650
Qy      661 SSPALXYLSPRTVLVSFVREELGVKARRGDVFQKQEVTTGSNVSRIYEAIKSGRLNXY 720
Db      651 SSPALXYLSPRTVLVSFVREELGVKARRGDVFQKQEVTTGSNVSRIYEAIKSGRLNXY 710
Qy      721 LVKMLA 726
Db      711 LKXMLA 716

RESULT 10
US-09-765-873A-37
; Sequence 37, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIORITY FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 2000-07-27
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X= Val, Met, Leu, Cys
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-37

Query Match      89.5%; Score 3010; DB 4; Length 716;

```


US-09-765-873A-33
; Sequence 33, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147, 719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X= Gly, Ser, Ala, Thr
; NAME/KEY: UNSURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
; NAME/KEY: UNSURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X= Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X= Val, Met, Leu, Cys
; NAME/KEY: UNSURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-33

Query Match 88.5%; Score 2978; DB 4; Length 716;
Best Local Similarity 82.9%; Pred. No. 0;
Matches 602; Conservative 25; Mismatches 89; Indels 10; Gaps 3;

QY 1 MAPSLSIATXSGXNGXGHAAXXASXXXXXXAAGSXLPTTXXTQLDIYEXXLD 60
DB 1 MAPSLSISHSPFANGASAKQAVNGAS-----TTLAAGSHLPTTQVTDIYEXKLLA 54
QY 61 PXTDDXXELDGYSLTLGDVVGAAARKGRVYVXDSEIRKXIDKSVFELRQXLNXYGV 120
DB 55 P-TDSTLELDGYSLNIGDVVSAARKGRPVYVXDSEIRKIDKSVFELRQXLSWYGV 113
QY 121 TGGGSAADRTEDALSLQALLLEHQLCGLPTSKXSFLGRLGLENLPLEVVRGAMTIRV 180
DB 114 TGGGSAADRTEDALSLQALLLEHQLCGLPSSFDXFRLLGRLGLENLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPVPFRGTISASGDLSPSYIAAITGHPDSKV 240
DB 174 NSLTRGSAVRLVLEALTNFLNHGITPVPFRGTISASGDLSPSYIAAISHGHPDSKV 233
QY 241 HXVHEXKIMAREALFGLPEVVLGPKSGLGIVNGTAVSASMATLALHDAMSLLS 300
DB 234 HXVHEXKILYAREAMALEFVVLGPKSGLGIVNGTAVSASMATLALHDAMSLLS 293

QY 301 QALTATVEAMVAGSFPFLHDVTRPHPHPTQIEVARNIRTLLEGSAFVHHEEVKXD 360
DB 294 QSLTAMTVEAMVAGSFPFLHDVTRPHPHPTQIEVAGNINKLEGSFANHHEEVKXD 353
QY 361 DEGIIRQDRYLRTPSPQMLGPLVSDMTAAHAYVLSLGGSTTNPPLIDENKXTHHGKGF 420
DB 354 DEGIIRQDRYLRTPSPQMLGPLVSDMTAAHAYVLSLGGSTTNPPLIDENKXTHHGKGF 413
QY 421 QASAYXNMTKTRIALLLIGKLNFTOLTENLNGMRGLSCLAEDPSLSYCKGLDJA 480
DB 414 QASAYXNMTKTRIALLLIGKLNFTOLTENLNGMRGLSCLAEDPSLSYCKGLDJA 473
QY 481 AAAYTSELGHLANPTTHVCPAENGQAVNSLLASRRTAEANDVLSLLATHLYCVIQ 540
DB 474 AAAYTSELGHLANPTTHVCPAENGQAVNSLLASRRTAEANDVLSLLATHLYCVIQ 533
QY 541 AYDLPAWEEFKQFPPKXXLLXQHRGXXXTXKXXELXKXVXKXKXKXKXKXKXKX 600
DB 534 AYDLPAWEEFKQFPPKXXLLXQHRGXXXTXKXXELXKXVXKXKXKXKXKXKXKX 593
QY 601 EPRWDAFSXATGTVEXLSSXXXXXVSLAANAMKVAAXEKAISLTXVXXFXWAPSS 660
DB 594 EPRWDAFSXATGTVEXLSSXXXXXVSLAANAMKVAAXEKAISLTXVXXFXWAPSS 650
QY 661 SSPALXYLSPRTVLYSFVREELGVKARQDVFLGKQVTTGNSVSRITFALSGKINXY 720
DB 651 SSPALXYLSPRTVLYSFVREELGVKARQDVFLGKQVTTGNSVSRITFALSGKINXY 710
QY 721 LYKMLA 726
DB 711 LYKMLA 716

RESULT 13
US-09-765-873A-32
; Sequence 32, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147, 719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 686
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-32

Query Match 88.3%; Score 2971; DB 4; Length 686;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 595; Conservative 23; Mismatches 68; Indels 4; Gaps 2;

QY 37 AAGSXLPTTXXTQLDIYEXXLDPTDXXELDGYSLTLGDVVGAAARKGRVYVXDSE 96
DB 1 AAGSXLPTTQVTDIYEXKLLA-PDSTLELDGYSLNIGDVVSAARKGRPVYVXDSE 59
QY 97 IRXKIDKSVFELRQXLNXYGVTTGGGSAADRTEDALSLQALLLEHQLCGLPTSKXS 156
DB 60 IRXKIDKSVFELRQXLSWYGVTTGGGSAADRTEDALSLQALLLEHQLCGLPSSFDS 119
QY 157 FYLGRGLENSLPLEVVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITPVPFRGTI 216
DB 120 FYLGRGLENSLPLEVVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITPVPFRGTI 179
QY 217 SASGDLSPSYIAAITGHPDSKVHXVHEXKIMAREALFGLPEVVLGPKSGLGIV 276
DB 204 SASGDLSPSYIAAITGHPDSKVHXVHEXKIMAREALFGLPEVVLGPKSGLGIV 266

```
Db 180 SASGDISPLSYIAAASGHPDSKVHVHEGKEKILYAREMAMFLNLEBPVVLGPKEGLGV 239
QY 277 NGTAVASAMATLALDPAHMLSLISQALATLVEAMVGHAGSFPHLQVTPREPTQIEVA 336
Db 240 NGTAVASAMATLALDPAHMLSLISQALATLVEAMVGHAGSFPHLQVTPREPTQIEVA 299
QY 337 RNIRLLSSGKFAVHHEEVKVKDEGLRQDRYPLRTSPQMLGPLVSDMTAHAVLSLE 396
Db 300 GNIRKLSSGKFAVHHEEVKVKDEGLRQDRYPLRTSPQMLGPLVSDMTAHAVLSLE 359
QY 397 AGOSTDNPLIVENKXTHHGNFQASAVXNMETRLALIGLNTQTQITENLNGMN 456
Db 360 AGOSTDNPLIVENKXTHHGNFQASAVXNMETRLALIGLNTQTQITENLNGMN 419
QY 457 RGLPSCIAAEDPSLSYHCKGLDIAAAAYTSELGLANPVTHVQPAEMGNQAVNSLALIS 516
Db 420 RGLPSCIAAEDPSLSYHCKGLDIAAAAYTSELGLANPVTHVQPAEMGNQAVNSLALIS 479
QY 517 ARTAEANVLSLLATHLVYCVLQAVDLPAHEFEKQFQXKXKXLLKHQFGKXKXKX 576
Db 480 ARTTESNDVLSLLATHLVYCVLQAVDLPAHEFEKQFQXKXKXLLKHQFGKXKXKX 539
QY 577 XXELKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 636
Db 540 RDELVEKNVKTAKLEQCNSTYDLPRKHDAFSFPAAGTIVEVLS--TSLSLAANVAMK 596
QY 637 VAXAEKALISLTXVXXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 606
Db 597 VAAAEKALISLTXVXXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 566
QY 697 QEVTTGSNVRIVEAIKSGRINXVLVKMLA 726
Db 657 QEVTTGSNVRIVEAIKSGRINXVLVKMLA 686

RESULT 14
US-09-624-693A-17
; Sequence 17, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocestra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Rhodotorula mucilaginosa
US-09-624-693A-17

Query Match 86.8%; Score 2922.5; DB 4; Length 713;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 595; Conservative 33; Mismatches 85; Indels 13; Gaps 5;

QY 1 MAPSLDISATSYXANGXNGXHAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
Db 1 MAPSVDSIATSYVANSISNGLHAAALAN-GSDVHKKTAGAGSLPTTETQTDIYERLAD 59
QY 61 PYTDXXELDGYSLTGLGVYGAARKGRVYVXDSDELTKKIDKSVBELRQOLXNSYVGT 120
Db 60 AGATDQIKLDGYTLTGLGVYGAARKGRVYVXDSDELTKKIDKSVBELRQOLXNSYVGT 119
QY 121 TQFGGADRTEDATSLQKALLEHQLCGVLPFSXDSFYLGSGLENSPLFVVRGAMTIRV 180
Db 120 TQFGGADRTEDATSLQKALLEHQLCGVLPFSXDSFYLGSGLENSPLFVVRGAMTIRV 179
QY 181 NSLIRGHSANVRLVLEALTNFLNHGITPVLPLRGTISASGDLSPLSYIAAITGHPSKV 240
```

```
Db 180 NSLIRGHSANVRLVLEALTNFLNHGITPVLPLRGTISASGDLSPLSYIAAITGHPSKV 239
QY 241 HYHXGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 300
Db 240 HV--DG--KIMSQAIALKGLQPVVLGPKEGLVNGTAVASAMATLALDPAHMLSL 295
QY 301 QALTALVEAMVGHAGSFPHLQVTPREPTQIEVARNIRITLSSGKFAVHHEEVKVKD 360
Db 296 QALTALVEAMVGHAGSFPHLQVTPREPTQIEVARNIRITLSSGKFAVHHEEVKVKD 355
QY 361 DEGIIRQDRYPLRTSPQMLGPLVSDMTAHAVLSLEAGQSTTDNPLIDVENKXTHHGNF 420
Db 356 DEGIIRQDRYPLRTSPQMLGPLVSDMTAHAVLSLEAGQSTTDNPLIDVENKXTHHGNF 415
QY 421 QASAVXNMETRLALALIGLNTQTQITENLNGMNGLPSCIAAEDPSLSYHCKGLDIA 480
Db 416 MASSVGNVTEKTRALVAMKVSFTQITENLNGMNGLPSCIAAEDPSLSYHCKGLDIA 475
QY 481 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARTAEANVLSLLATHLVYCVLQ 540
Db 476 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARTAEANVLSLLATHLVYCVLQ 535
QY 541 AYDLPAHEFEKQFQXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db 536 AYDLPAHEFEKQFQXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 590
QY 601 EPRMHDAFSXATGTVEEXLSXXXXXVSLAANVAMKVAAXAEKALISLTXVXXKXKX 660
Db 591 EQRHMDTSVATGAVVELAG---QEVSLASLNAMKVAACEKALALRVSRRDSFWAAPS 647
QY 661 SSPALXYLSPTRVLYSFVREELGVKARGDVLGKQEVTTGSNVRIVEAIKSGRINXV 720
Db 648 SSPALXYLSPTRVLYSFVREELGVKARGDVLGKQEVTTGSNVRIVEAIKSGCIAPV 707
QY 721 LYKMLA 726
Db 708 LYKMLA 713

RESULT 15
US-09-624-693A-13
; Sequence 13, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocestra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Rhodotorula graminis
US-09-624-693A-13

Query Match 82.3%; Score 2770; DB 4; Length 720;
Best Local Similarity 77.5%; Pred. No. 4,6e-309;
Matches 563; Conservative 40; Mismatches 117; Indels 6; Gaps 3;

QY 1 MAPSLDISATSYXANGXNGXHAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
Db 1 MAPSLDISATSYXANGXNGXHAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
QY 61 PYTDXXELDGYSLTGLGVYGAARKGRVYVXDSDELTKKIDKSVBELRQOLXNSYVGT 120
Db 61 PYTDXXELDGYSLTGLGVYGAARKGRVYVXDSDELTKKIDKSVBELRQOLXNSYVGT 119
```

QY 121 TGFSGSADTRTEDAISLOKALLEHQLCGVLPISXDSFXLGRGLENSLPLEVVRGAMTIRV 180
 Db 120 TGFSGSADTRTEDAVALOKALIEHQLCGVLPISVSSFSVGRGLENTLPLEVVRGAMTIRV 179
 QY 181 NSLIRGHSARLVVLEALTNFNLHGTPVPLRGITISASCDLSPISYIAAITGHPDSKY 240
 Db 180 NSLIRGHSARLVLEALTNFNLHGTPVPLRGITISASCDLSPISYIAGITGHPDKV 239
 QY 241 HVHGEKEXIMXAREALFGLPEVYLGPKEGLVNGTAVASAMATLADAHMLSL 300
 Db 240 HVHGEKEXIMFAREALSLFGLPAVVLGPKEGLVNGTAVASAMATLADAHMLSL 299
 QY 301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLBSSKFAVHHEEVKXD 360
 Db 300 QALTALTVEAMVQCGSFAPFIHDYCRPHGQVEVARNIRTLSSGSFAVEHEEVKXD 359
 QY 361 DEGILRODRPLRTSPQWIGPLVSDMIAHNAVLSLEAGOSTDNPDIIDVENKXTHHGNF 420
 Db 360 DEGILRODRPLRTSPQWIGPLVEDMMAVSTLSLE -NNTTNDPDIIDVENKQTAHGNF 418
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
 Db 419 QASAVSISMEKTRIALALIGKLNFTQCTELINAAANNRGLPSCLAEDPSLNYHCKGLDIH 478
 QY 481 AAAYTSEIGHLANPVTTHVOPAEKNOAVNSLALISARRTAEANDVLSLLATHLVCVLO 540
 Db 479 IAAVASSELGHLANPVTTFVOPAEKNOAVNSLALISARRTAEANDVLSLLASHLYCTLQ 538
 QY 541 AVDLRAMEFEFFKQFPXXXXLLXQHFXXXXTXXXXXLXXKVKXLXKRLBQDNSYDL 600
 Db 539 AVDLRAMELDFKQFPDPLPTLLQOHLGTGLDVNA---LALBVKKALNKRLBQDTTYDL 594
 QY 601 EPRMHDAFSXATGTVEKLSXXXXXVSLAANANKVAXAEKALSLTRVXXXFWXAPSS 660
 Db 595 EPRMHDAFSYATGTVEKLSPPSANVTLTAVANAKVASAEKALSLTREVRNRFWQTPSS 654
 QY 661 SSPALXYLSPTREVLXSFVREELGVARRGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 Db 655 QAPAHAYLSPTREVLXSFVREELGVARRGDVFGVQOETIGSNVSRITYEAIKQGRINHV 714
 QY 721 LVFMALA 726
 Db 715 LVFMALA 720

Search completed: September 9, 2004, 10:09:41
 Job time : 26.5975 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:07:54 ; Search time 93.3859 Seconds
(without alignments)
2493.103 Million cell updates/sec

Title: US-09-939-408a-21
Perfect score: 3365
Sequence: 1 MAPSDSLATXKXNGXNGX.....RIYAKSGRINXVYKMLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 32068961 residues
Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US00_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3323	95.8	726	US-09-939-408a-21	Sequence 2, Appl
2	3021	89.8	716	US-09-765-873a-8	Sequence 8, Appl
3	3021	89.8	716	US-09-939-408a-19	Sequence 19, Appl
4	3021	89.8	716	US-10-138-970a-4	Sequence 4, Appl
5	3021	89.8	716	US-10-168-523b-8	Sequence 8, Appl
6	3021	89.8	716	US-10-374-366-2	Sequence 2, Appl
7	3021	89.8	716	US-10-439-478-2	Sequence 2, Appl
8	3019	89.7	716	US-09-765-873a-10	Sequence 10, Appl
9	3019	89.7	716	US-10-138-970a-24	Sequence 24, Appl
10	3019	89.7	716	US-10-168-523b-10	Sequence 10, Appl
11	3019	89.7	716	US-10-374-366-4	Sequence 4, Appl
12	3016	89.6	716	US-10-374-366-24	Sequence 24, Appl
13	3015	89.6	716	US-10-374-366-22	Sequence 22, Appl
14	3014	89.6	716	US-09-765-873a-38	Sequence 38, Appl
15	3014	89.6	716	US-10-138-970a-22	Sequence 22, Appl

16	3014	89.6	716	US-10-168-523b-38	Sequence 38, Appl
17	3014	89.6	716	US-10-374-366-16	Sequence 16, Appl
18	3014	89.6	716	US-10-439-478-18	Sequence 18, Appl
19	3013	89.5	716	US-10-374-366-8	Sequence 8, Appl
20	3012	89.5	716	US-09-765-873a-35	Sequence 35, Appl
21	3012	89.5	716	US-10-138-970a-19	Sequence 19, Appl
22	3012	89.5	716	US-10-168-523b-35	Sequence 35, Appl
23	3012	89.5	716	US-10-374-366-10	Sequence 10, Appl
24	3012	89.5	716	US-10-439-478-15	Sequence 15, Appl
25	3011	89.5	716	US-10-374-366-14	Sequence 14, Appl
26	3010	89.5	716	US-09-765-873a-36	Sequence 36, Appl
27	3010	89.5	716	US-09-765-873a-37	Sequence 37, Appl
28	3010	89.5	716	US-10-138-970a-20	Sequence 20, Appl
29	3010	89.5	716	US-10-138-970a-21	Sequence 21, Appl
30	3010	89.5	716	US-10-168-523b-36	Sequence 36, Appl
31	3010	89.5	716	US-10-168-523b-37	Sequence 37, Appl
32	3010	89.5	716	US-10-439-478-16	Sequence 16, Appl
33	3010	89.5	716	US-10-439-478-17	Sequence 17, Appl
34	3006	89.3	716	US-09-765-873a-34	Sequence 34, Appl
35	3006	89.3	716	US-10-138-970a-18	Sequence 18, Appl
36	3006	89.3	716	US-10-168-523b-34	Sequence 34, Appl
37	3006	89.3	716	US-10-374-366-18	Sequence 18, Appl
38	3006	89.3	716	US-10-439-478-14	Sequence 14, Appl
39	3005	89.3	716	US-10-374-366-6	Sequence 6, Appl
40	3005	89.3	716	US-10-374-366-20	Sequence 20, Appl
41	2997	89.1	716	US-10-374-366-12	Sequence 12, Appl
42	2978	88.5	716	US-09-765-873a-33	Sequence 33, Appl
43	2978	88.5	716	US-10-138-970a-17	Sequence 17, Appl
44	2978	88.5	716	US-10-168-523b-33	Sequence 33, Appl
45	2978	88.5	716	US-10-439-478-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-939-408a-21
; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939, 408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624, 693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Yaa"means any amino acid; "Xaa"means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
US-09-939-408a-21

Query Match 95.8%; Score 3323; DB 9; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPSDSLATXKXNGXNGXAAAXXXXXXXXXXXXXXAGSLPTTXXTQDLIVEXXIAD 60
1 MAPSDSLATXKXNGXNGXAAAXXXXXXXXXXXXXXAGSLPTTXXTQDLIVEXXIAD 60
Db 1 MAPSDSLATXKXNGXNGXAAAXXXXXXXXXXXXXXAGSLPTTXXTQDLIVEXXIAD 60

Query 61 PXTDDXXELDGYSLTIGDVVGAAKRGKRYRVXDSPDEIRKIDKSVFELRXOLXNSVYCVT 120
 Db 61 PXTDDXXELDGYSLTIGDVVGAAKRGKRYRVXDSPDEIRKIDKSVFELRXOLXNSVYCVT 120
 QY 121 TGFSGSADRTEDAIISLQKALLEHQLCVLPISXSPXFLGRLGLNSLPLEVVRGAMTIRV 180
 Db 121 TGFSGSADRTEDAIISLQKALLEHQLCVLPISXSPXFLGRLGLNSLPLEVVRGAMTIRV 180
 QY 181 NSLTRGHSARLVVLEALTNFLNHGTPVPRGRTISASGDLSPSTYIAAITHGHPDSKY 240
 Db 181 NSLTRGHSARLVVLEALTNFLNHGTPVPRGRTISASGDLSPSTYIAAITHGHPDSKY 240
 QY 241 HYNHEGEXEKIMKAREAIALFGLPEPVYLGPKEGGLVNGTAVASAMATLALDHAMLSLS 300
 Db 241 HYNHEGEXEKIMKAREAIALFGLPEPVYLGPKEGGLVNGTAVASAMATLALDHAMLSLS 300
 QY 301 QALTALTVTEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 Db 301 QALTALTVTEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 QY 361 DEGLRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGGSTTDNPLIDVENKXTHHGNF 420
 Db 361 DEGLRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGGSTTDNPLIDVENKXTHHGNF 420
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNMGRLPSCLAEDPSSLSTHCKGLDIA 480
 Db 421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNMGRLPSCLAEDPSSLSTHCKGLDIA 480
 QY 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLCVLO 540
 Db 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLCVLO 540
 QY 541 AVDLRAMEFEFKQFPXKXLLXQHFQXXXTXXXXXXELXXKXKXKXLEQTNSTYDL 600
 Db 541 AVDLRAMEFEFKQFPXKXLLXQHFQXXXTXXXXXXELXXKXKXKXLEQTNSTYDL 600
 QY 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAANAKVAXAEKASISLTRVXXKXFWXAPS 660
 Db 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAANAKVAXAEKASISLTRVXXKXFWXAPS 660
 QY 661 SSPALXYLSPRTVLSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 Db 661 SSPALXYLSPRTVLSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 QY 721 LVKOLA 726
 Db 721 LVKOLA 726

RESULT 2
 US-09-765-873A-8
 ; Sequence 8, Application US/09765873A
 ; Patent No. US20010053847A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-Song
 ; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
 ; FILE REFERENCE: BC1009 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/765, 873A
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: US 09/627, 216
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR APPLICATION NUMBER: US 60/147, 719
 ; PRIOR FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 8
 ; LENGTH: 716
 ; TYPE: PRT
 ; ORGANISM: Rhodocorula glutinis
 US-09-765-873A-8
 Query Match 89.8%; Score 3021; DB 9; Length 716;

Best Local Similarity 84.0%; Pred. No. 7,1e-298;
 Matches 610; Conservatve 25; Mismatches 81; Indels 10; Gaps 3;
 QY 1 MAPSIDSTATSANGKXNGKHAAXXAXXXXXXAGSKLPTTYXTOLDIVEKXAD 60
 Db 1 MAPSIDSTATSANGKXNGKHAAXXAXXXXXXAGSKLPTTYXTOLDIVEKXAD 60
 QY 1 MAPSIDSTATSANGKXNGKHAAXXAXXXXXXAGSKLPTTYXTOLDIVEKXAD 60
 Db 1 MAPSIDSTATSANGKXNGKHAAXXAXXXXXXAGSKLPTTYXTOLDIVEKXAD 60
 QY 55 P-TDSTLBDGYSLNLDGVVGAARKGRPVYKXDSDEIRKIDKSVFELRXOLXNSVYCVT 113
 Db 55 P-TDSTLBDGYSLNLDGVVGAARKGRPVYKXDSDEIRKIDKSVFELRXOLXNSVYCVT 113
 QY 121 TGFSGSADRTEDAIISLQKALLEHQLCVLPISXSPXFLGRLGLNSLPLEVVRGAMTIRV 180
 Db 121 TGFSGSADRTEDAIISLQKALLEHQLCVLPISXSPXFLGRLGLNSLPLEVVRGAMTIRV 180
 QY 181 NSLTRGHSARLVVLEALTNFLNHGTPVPRGRTISASGDLSPSTYIAAITHGHPDSKY 240
 Db 181 NSLTRGHSARLVVLEALTNFLNHGTPVPRGRTISASGDLSPSTYIAAITHGHPDSKY 240
 QY 241 HYNHEGEXEKIMKAREAIALFGLPEPVYLGPKEGGLVNGTAVASAMATLALDHAMLSLS 300
 Db 241 HYNHEGEXEKIMKAREAIALFGLPEPVYLGPKEGGLVNGTAVASAMATLALDHAMLSLS 300
 QY 301 QALTALTVTEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 Db 301 QALTALTVTEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 QY 361 DEGLRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGGSTTDNPLIDVENKXTHHGNF 420
 Db 361 DEGLRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGGSTTDNPLIDVENKXTHHGNF 420
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNMGRLPSCLAEDPSSLSTHCKGLDIA 480
 Db 421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNMGRLPSCLAEDPSSLSTHCKGLDIA 480
 QY 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLCVLO 540
 Db 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLCVLO 540
 QY 541 AVDLRAMEFEFKQFPXKXLLXQHFQXXXTXXXXXXELXXKXKXKXLEQTNSTYDL 600
 Db 541 AVDLRAMEFEFKQFPXKXLLXQHFQXXXTXXXXXXELXXKXKXKXLEQTNSTYDL 600
 QY 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAANAKVAXAEKASISLTRVXXKXFWXAPS 660
 Db 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAANAKVAXAEKASISLTRVXXKXFWXAPS 660
 QY 661 SSPALXYLSPRTVLSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 Db 661 SSPALXYLSPRTVLSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 QY 721 LVKOLA 726
 Db 721 LVKOLA 726

RESULT 3
 US-09-939-408A-19
 ; Sequence 19, Application US/09939408A
 ; Patent No. US20020102712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yoshida, Roberta
 ; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 ; FILE REFERENCE: 29479/500NSCA
 ; CURRENT APPLICATION NUMBER: US/09/939, 408A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 09/624, 693
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: PCT/US01/23270
 ; PRIOR FILING DATE: 2001-07-24


```
Db 651 SSPALSYSPRTQILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKTYEALISGRINNV 710
QY 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 7
US-10-439-478-2
; Sequence 2, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Basat, Arle
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Vanelli, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxyetylene
; FILE REFERENCE: C11912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-439-478-2

Query Match      89.8%; Score 3021; DB 15; Length 716;
Best Local Similarity 84.0%; Pred. No. 7,1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSDSIATSXANGXNGHAAAXXAXXXXXXAXAGXLPPTXXTQLDIVEXXLD 60
Db 1 MAPSDSIHSHFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTVQVDIVEXKMLA 54
QY 61 PXTDXKXELDGSILTLGDVYGAARKGRVYXDSDEIRKXIKDKSVEFLRQOLXNSVYGT 120
Db 55 P-TDSTLELDGSLNLGDVSAARKGRVYXDSDEIRSKIKDKSVEFLRQOLXNSVYGT 113
QY 121 TGFSGSADTRTEDAISILOKALLEHQLCGVLPTSXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISILOKALLEHQLCGVLPSDFSFLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 240
Db 174 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 233
QY 241 HVYHEGEXKIMAREAILFGLPEVVLGPKGELGVNGTAVASAMATLALHDAHMLSLLS 300
Db 234 HVYHEGEXKILYAREKMALENLEPVVLGPKGELGVNGTAVASAMATLALHDAHMLSLLS 293
QY 301 QALTLATVEAMVGAHGSFHPFLADVTRPHPQIEVANIRITLLEGSXFAYVHHEEVKVD 360
Db 294 QSLTAMTVEAMVGAHGSFHPFLADVTRPHPQIEVANIRITLLEGSFAYVHHEEVKVD 353
QY 361 DSGILRQDYPRLRTPSQWGLPVSDMHAHVLISLEAGOSTTNDPLDVENKXTHHGNF 420
Db 354 DSGILRQDYPRLRTPSQWGLPVSDMHAHVLISLEAGOSTTNDPLDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKINFOTLEMLNAGNRLPSCLAADPSSLSYHCKGLDIA 480
Db 414 QASAVXNTEKTRIRLGAQIGKINFOTLEMLNAGNRLPSCLAADPSSLSYHCKGLDIA 473
QY 481 AAAYSELCHLANPYTTTHVQPAEMGNQAVNSIALISARITLAVNDVLSLIATILYCVLQ 540
Db 474 AAAYSELCHLANPYTTTHVQPAEMGNQAVNSIALISARITLAVNDVLSLIATILYCVLQ 533
```

```
QY 541 AYDLRAMEFEFKQFPKXXXXLLXQHFQXXXTXXXXXXELXKXKXKXKXLEQTNVDL 600
Db 534 AIDLRIIEFEFKQFGPAIVSLIDQHFQSANTGSLNDELVEKXKXKXKXLEQTNVDL 593
QY 601 EPRMHDASFAXATGTVEXXSSXXXXXVSLAAVNAKMYAXAEKASITLTRYXKXFWKAPSS 660
Db 594 VPRMHDASFAXATGTVEXXSS-----TSLSLAAVNAKMYAAAEASISLRQVRETFWMAAST 650
QY 661 SSPALXYISPRTRVLYSVREELGVKARRGDVFLGKQEVITIGSNVSKTYEALISGRINNV 720
Db 651 SSPALSYSPRTQILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKTYEALISGRINNV 710
QY 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 8
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-10

Query Match      89.7%; Score 3019; DB 9; Length 716;
Best Local Similarity 84.0%; Pred. No. 1,1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MAPSDSIATSXANGXNGHAAAXXAXXXXXXAXAGXLPPTXXTQLDIVEXXLD 60
Db 1 MAPSDSIHSHFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTVQVDIVEXKMLA 54
QY 61 PXTDXKXELDGSILTLGDVYGAARKGRVYXDSDEIRKXIKDKSVEFLRQOLXNSVYGT 120
Db 55 P-TDSTLELDGSLNLGDVSAARKGRVYXDSDEIRSKIKDKSVEFLRQOLXNSVYGT 113
QY 121 TGFSGSADTRTEDAISILOKALLEHQLCGVLPTSXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISILOKALLEHQLCGVLPSDFSFLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 240
Db 174 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 233
QY 241 HVYHEGEXKIMAREAILFGLPEVVLGPKGELGVNGTAVASAMATLALHDAHMLSLLS 300
Db 234 HVYHEGEXKILYAREKMALENLEPVVLGPKGELGVNGTAVASAMATLALHDAHMLSLLS 293
QY 301 QALTLATVEAMVGAHGSFHPFLADVTRPHPQIEVANIRITLLEGSXFAYVHHEEVKVD 360
Db 294 QSLTAMTVEAMVGAHGSFHPFLADVTRPHPQIEVANIRITLLEGSFAYVHHEEVKVD 353
QY 361 DSGILRQDYPRLRTPSQWGLPVSDMHAHVLISLEAGOSTTNDPLDVENKXTHHGNF 420
Db 354 DSGILRQDYPRLRTPSQWGLPVSDMHAHVLISLEAGOSTTNDPLDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKINFOTLEMLNAGNRLPSCLAADPSSLSYHCKGLDIA 480
Db 414 QASAVXNTEKTRIRLGAQIGKINFOTLEMLNAGNRLPSCLAADPSSLSYHCKGLDIA 473
```

Db 414 QAAVAANTMEKTRGLAIGKLNFTQITEMLNAQNRGLPSCLAAEDPSLSYHCKGLDIA 473
Qy 481 AAATSEGLHANLPVTHVQPAEMGNQAVNSLALISARTAEANDVLSLLATHLYCYLQ 540
Db 474 AAATSEGLHANLPVTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLYCYLQ 533
Qy 541 AVDLRAMEFEFKQFPAXXXXLLXOHFGXXXXTXXXKXXKXKXLEQTNXYDL 600
Db 534 AIDLRATEFEFKQFGPAIVSLIDOHFGSAMTGSNLRDELVEKNKTLAKLEQTNXYDL 593
Qy 601 EPRWHDAPSAATGTVVEVLSXXXXXVSLAANAMKVAAXEKAISLTREYXXFXAPSS 660
Db 594 VPRWHDAPSAATGTVVEVLS--TSLSLAAVNAWKVAABESAI SLTRQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLYSFVREELGVKARPGDVFGLKQEVITGSNVSRIYEAIRSGRINXV 720
Db 651 SSPALXYLSPTRTQILYAFVREELGVKARPGDVFGLKQEVITGSNVSRIYEAIRSGRINNV 710
Qy 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 9
US-10-138-970A-24
; Sequence 24, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; NUMBER OF FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; LENGTH: 716
; TYPE: PRF
; ORGANISM: artificial sequence
; OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24

Query Match 89.7%; Score 3019; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 1.1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

Qy 1 MAPSLDSIATSKXNGXKXGHAAXXASXXXXXXXAGSKLPTTXXTQDIVEXXLAD 60
Db 1 MAPSLDSISHSFANGVASAKQAVNGAS-----TNLAAVAGSHLPTTQVTVQDIVEXXLAA 54
Qy 61 PYTDDXXELDGYSLTLGADVGAARKGRXVRVXDSEIRXKIDKSVEFRLXQLNXYGYVT 120
Db 55 P-TDSTLELDGYSLNLDGVVSAARKGRPVRYKDSDEIRSKIDKSVEFRLSOLMSYGYVT 113
Qy 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENSLPLEVVRGAMTIRV 173
Qy 181 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISASGDLSPISYIAAITGHPDSKY 240
Db 174 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISASGDLSPISYIAAISGHPDSKY 233
Qy 241 HVXHEGEXIKMAREALFGLBPVVLGPKEGGLGVNGTAVASAMATLADHMTLSLS 300
Db 234 HVXHEGEXIKLVARBAMALFNLBPVVLGPKEGGLGVNGTAVASAMATLADHMTLSLS 293
Qy 301 QALTALTAVAMVGHASPHRLADVTYRPHPTQIEVARYTRILLSSYAFVHHEEYVKD 360
Db 294 QSLTAMTAVAMVGHASPHRLADVTYRPHPTQIEVAGIRIKLLSRRAPVHHEEYVKD 353

Qy 361 DEGILRODRYPLRTSPQWLGPLVSDMTAAHAVLSLEAGOSTTNDPLIDVENKXTHHGGNF 420
Db 354 DEGILRODRYPLRTSPQWLGPLVSDMTAAHAVLSLEAGOSTTNDPLIDVENKXTHHGGNF 413
Qy 421 QASAVXANTMEKTRGLAIGKLNFTQITEMLNAQNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db 414 QAAVAANTMEKTRGLAIGKLNFTQITEMLNAQNRGLPSCLAAEDPSLSYHCKGLDIA 473
Qy 481 AAATSEGLHANLPVTHVQPAEMGNQAVNSLALISARTAEANDVLSLLATHLYCYLQ 540
Db 474 AAATSEGLHANLPVTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLYCYLQ 533
Qy 541 AVDLRAMEFEFKQFPAXXXXLLXOHFGXXXXTXXXKXXKXKXLEQTNXYDL 600
Db 534 AIDLRATEFEFKQFGPAIVSLIDOHFGSAMTGSNLRDELVEKNKTLAKLEQTNXYDL 593
Qy 601 EPRWHDAPSAATGTVVEVLSXXXXXVSLAANAMKVAAXEKAISLTREYXXFXAPSS 660
Db 594 VPRWHDAPSAATGTVVEVLS--TSLSLAAVNAWKVAABESAI SLTRQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLYSFVREELGVKARPGDVFGLKQEVITGSNVSRIYEAIRSGRINXV 720
Db 651 SSPALXYLSPTRTQILYAFVREELGVKARPGDVFGLKQEVITGSNVSRIYEAIRSGRINNV 710
Qy 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 10
US-10-188-523B-10
; Sequence 10, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; LENGTH: 716
; TYPE: PRF
; ORGANISM: Rhodotorula glutinis mutant
US-10-188-523B-10

Query Match 89.7%; Score 3019; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 1.1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

Qy 1 MAPSLDSIATSKXNGXKXGHAAXXASXXXXXXXAGSKLPTTXXTQDIVEXXLAD 60
Db 1 MAPSLDSISHSFANGVASAKQAVNGAS-----TNLAAVAGSHLPTTQVTVQDIVEXXLAA 54
Qy 61 PYTDDXXELDGYSLTLGADVGAARKGRXVRVXDSEIRXKIDKSVEFRLXQLNXYGYVT 120
Db 55 P-TDSTLELDGYSLNLDGVVSAARKGRPVRYKDSDEIRSKIDKSVEFRLSOLMSYGYVT 113
Qy 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENSLPLEVVRGAMTIRV 173
Qy 181 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISASGDLSPISYIAAITGHPDSKY 240
Db 174 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISASGDLSPISYIAAISGHPDSKY 233
Qy 241 HVXHEGEXIKMAREALFGLBPVVLGPKEGGLGVNGTAVASAMATLADHMTLSLS 300

```
Db      234 HVHGEKKEKILYARBAALFMLEPVLGPKGGLGVNGTAVASAMATLADHADMILSLLS 293
QY      301 QALTAITYEAMVGHAGSFHPFLHDVTRPHTQIEVARNIRTLLEGXFAVHHEEVVKD 360
Db      294 QSLTMTVEAMVGHAGSFHPFLHDVTRPHTQIEVARNIRTLLEGXFAVHHEEVVKD 353
QY      361 DEGIIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 420
Db      354 DEGIIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 413
QY      421 QASAVXNTMEKTRIALALIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 480
Db      414 QAAAVANTMEKTRIALALIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 473
QY      481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLATHLVCVLQ 540
Db      474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAESNDVLSLLATHLVCVLQ 533
QY      541 AVDLRAMEFEFKQFPXKXXLIXQHFQXXXLXKXXKXXKXKXRLBQTSNYDL 600
Db      534 AIDLRAEFERKQFPXKXXLIXQHFQXXXLXKXXKXXKXKXRLBQTSNYDL 593
QY      601 EPRWMDAFSXAATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVYXXFWXAPSS 660
Db      594 VPRWMDAFSXAATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVYXXFWXAPSS 650
QY      661 SSPALXYISPRTRVLYSVFRELGVKARRGDVFLGKQEVITGSVNSRIYEAISGRINXY 720
Db      651 SSPALXYISPRTRVLYSVFRELGVKARRGDVFLGKQEVITGSVNSRIYEAISGRINXY 710
QY      721 LVKMLA 726
Db      711 LVKMLA 716

RESULT 11
US-10-374-366-4
; Sequence 4, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CIL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-4

Query Match      89.7%; Score 3019; DB 15; Length 716;
Best Local Similarity 84.0%; Pred. No. 1,1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;
```

```
QY      181 NSLTRGSAVRLVLEALTNFLNHGTTPIVPLRGTTISAGDLSPLSTIAAATGHPDSKY 240
Db      174 NSLTRGSAVRLVLEALTNFLNHGTTPIVPLRGTTISAGDLSPLSTIAAATGHPDSKY 223
QY      241 HVHGEKKEKIXABEALALFGLPEVVLGPKKGLGVNGTAVSAMATLADHADMILSLLS 300
Db      234 HVHGEKKEKILYARBAALFMLEPVLGPKKGLGVNGTAVSAMATLADHADMILSLLS 293
QY      301 QALTAITYEAMVGHAGSFHPFLHDVTRPHTQIEVARNIRTLLEGXFAVHHEEVVKD 360
Db      294 QSLTMTVEAMVGHAGSFHPFLHDVTRPHTQIEVARNIRTLLEGXFAVHHEEVVKD 353
QY      361 DEGIIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 420
Db      354 DEGIIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 413
QY      421 QASAVXNTMEKTRIALALIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 480
Db      414 QAAAVANTMEKTRIALALIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 473
QY      481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLATHLVCVLQ 540
Db      474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARRTAESNDVLSLLATHLVCVLQ 533
QY      541 AVDLRAMEFEFKQFPXKXXLIXQHFQXXXLXKXXKXXKXKXRLBQTSNYDL 600
Db      534 AIDLRAEFERKQFPXKXXLIXQHFQXXXLXKXXKXXKXKXRLBQTSNYDL 593
QY      601 EPRWMDAFSXAATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVYXXFWXAPSS 660
Db      594 VPRWMDAFSXAATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVYXXFWXAPSS 650
QY      661 SSPALXYISPRTRVLYSVFRELGVKARRGDVFLGKQEVITGSVNSRIYEAISGRINXY 720
Db      651 SSPALXYISPRTRVLYSVFRELGVKARRGDVFLGKQEVITGSVNSRIYEAISGRINXY 710
QY      721 LVKMLA 726
Db      711 LVKMLA 716

RESULT 12
US-10-374-366-24
; Sequence 24, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CIL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 24
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-24

Query Match      89.6%; Score 3016; DB 15; Length 716;
Best Local Similarity 83.9%; Pred. No. 2,3e-297;
Matches 609; Conservative 25; Mismatches 82; Indels 10; Gaps 3;
```

Page 8

```

Db      I  MBSLSDISHSPANGVASKQAVNGAS-----TMLAVAGSHLEPTTQVTVQVIVKRLAA 54
Qy      61  PXTDDXELLDGYSLTLDGVVGAARRGRXRVXVDSDEIRKIDKSEVFLRQJLANSYGV 120
Db      55  P-TDSTLELDGYSINLGDVVASAARKGRBVRVWDSDEIRSKIDKSEVFLRQJLSMSYGV 113
Qy      121  TGGGSADRTDASISLOKALLEHOLQGVLTFSXDSFPLSGGLENSIPLSEVGMAMTRV 160
Db      114  TGGGSADRTDASISLOKALLEHOLQGVLTFSXDSFPLSGGLENSIPLSEVGMAMTRV 173
Qy      181  NSLTRSHASAVRLVLEALTNFLNHGITIVPLRGTISASGDSLSELYIAAIIIGHDPDSKV 240
Db      174  NSLTRSHASAVRLVLEALTNFLNHGITIVPLRGTISASGDSLSELYIAAIIIGHDPDSKV 223
Qy      241  HXHXHEKEXIKMARERIALFGJLEPVVLPFGKGLVNGTVASAMATLALHDAHMSLS 300
Db      234  HXHXHEKEXIKLARERAMALFNLEPVVLPFGKGLVNGTVASAMATLALHDAHMSLS 293
Qy      301  QALTALTAVAMGHAAGSFHFLADVTREPHPOIEVARNIRTLBGSXFVAHHEEYVKD 360
Db      294  QSLTALTAVAMGHAAGSFHFLADVTREPHPOIEVAGNIRKLEBGSFPAHHEEYVKD 353
Qy      361  DEGLRQDRYPLRTSPQWLGPLVSDMIAHAAVLSLEAGQSTTNP.LIDENKXTHHGGNF 420
Db      354  DEGLRQDRYPLRTSPQWLGPLVSD.LIAHAAVLTLEAGQSTTNP.LIDENKTHSHGGNF 413
Qy      421  QSAVYVNMTEKRLALIGLINFQULTEMLNAGMNGLPSCLAEPSISYHCKGLDIA 480
Db      414  QSAVYVNMTEKRLGLAQIGKINFQULTEMLNAGMNGLPSCLAEPSISYHCKGLDIA 473
Qy      481  AAAYTSELGLNAPYTHVOPAEKMGQAVNSIALISARRTAEADVLSLLATHLVCVQ 540
Db      474  AAAYTSELGLNAPYTHVOPAEKMGQAVNSIALISARRTESNDVLSLLATHLVCVQ 533
Qy      541  AAVLRAMEFEFKQKXPXXXXLXKHQFXXXXLXXXXXELXXXXKXKXLRLEQTSYDL 600
Db      534  AAVLRAMEFEFKQKGPVLSLIDQHFSSAMTGSNLDLEVEKXNKTLARLQTSYDL 593
Qy      601  EPRMHDAPFAXATGVVEXKLSXXXXXVSLAVNAMKAXAAXEASLSTRXVXXKFXKAPSS 660
Db      594  VPRMHDAPFAGCTIVEVLSS---TSLSLAVNMKXALAAESASLTLQVRETFWMAAS 650
Qy      661  SSPALXYLSPRTVLYSFVREELGVKARGDVFJGQEVTTGSVNSIYPAIKSGINXY 720
Db      651  SSPALXYLSPRTQIILYAFVREELGVKARGDVFJGQEVTTGSVNSIYPAIKSGINXY 710
Qy      721  LVMKLA 726
Db      711  LLMKLA 716

RESULT 14
US-09-765-873A-38
; Sequence 38, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38

```

```

? SEQ ID NO 38
? LENGTH: 716
? TYPE: PRT
? ORGANISM: mutant from Rhodotorula glutinis
? FEATURE:
? NAME/KEY: UNSURE

```

LOCATION: (502) : (502)
OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
NAME/KEY: UNSURE
LOCATION: (540) : (540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match 89.6%; Score 3014; DB 9; Length 716;
Best Local Similarity 83.9%; Pred. No. 3,7e-297;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

```
QY 1 MAPSLDSTATSYXANGXNKGXHAAXXASAXXXXXXAGSXLPTTXXQTOLDIYEXKLAD 60
DB 1 MAPSLDSTSHSANGVASKQAVNGAS-----TNLVAAGSHLPTTQYQYQYDIYEXKLAA 54
QY 61 PXTDDXHELDGSLTLGQVGAARKGRVYXDSDEIRXKIDKSEVEFLRQLSMSYGV 120
DB 55 P-TDSTLELDGSLTLGQVGAARKGRVYXDSDEIRXKIDKSEVEFLRQLSMSYGV 113
QY 121 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSXDFLGLGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSXDFLGLGLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSHAVRLVLEALTNFLNHGITPVLKGTISASGDLSPLSYTAATGHPDSKY 240
DB 174 NSLTRGSHAVRLVLEALTNFLNHGITPVLKGTISASGDLSPLSYTAATGHPDSKY 233
QY 241 HVYHGEKXIKMAREBALFGLPEVVLGPKGGLGVNGTAVASAMATLALDHAMLSLS 300
DB 234 HVYHGEKXIKLAREBALFGLPEVVLGPKGGLGVNGTAVASAMATLALDHAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDYTRPHPTQIEVARNIRTLGSGXFAYHHEEYVKD 360
DB 294 QSLTAMTYEAMVGHAGSFHPLHDYTRPHPTQIEVAGNIRTLGSGRFAYHHEEYVKD 353
QY 361 DEGLRQDRYP-RTSPQWLGPLVSDMIAHAVLSLEAGOSTTNDPLLDVENKXTHHGGNF 420
DB 354 DEGLRQDRYP-RTSPQWLGPLVSDMIAHAVLSLEAGOSTTNDPLLDVENKXTHHGGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 480
DB 414 QAAAVANTMEKTRILAOIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 473
QY 481 AAAYTSELGHLANPYTHVOPAEWNGQAVNSIALISARRTAEANDVLSLLATLHYCVLQ 540
DB 474 AAAYTSELGHLANPYTHVOPAEWNGQAVNSIALISARRTESNDVLSLLATLHYCVLQ 533
QY 541 AYDLRAMEFEFKQFPPXXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTN SYDL 600
DB 534 AYDLRAMEFEFKQFPPXXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTN SYDL 593
QY 601 EPRMHDASFATGTVEXLSXXXXXVSLAAVNAWKVAXAEKASLTRYVXXFXMXPSS 660
DB 594 VERMHDASFATGTVEXLS-----TSLSLAAVNAWKVAXAEKASLTRYVXXFXMXPSS 650
QY 661 SSPALXYLSPTRRVLYSFVREELGVKARGDVFGLGQEVTTIGSVNSRIYEAIKSGRINXV 720
DB 651 SSPALXYLSPTRRVLYSFVREELGVKARGDVFGLGQEVTTIGSVNSRIYEAIKSGRINXV 710
QY 721 LVKMLA 726
DB 711 LKMLA 716
```

RESULT 15
US-10-138-970A-22
Sequence 22, Application US/10138970A
Publication No. US20030079255A1
GENERAL INFORMATION:
APPLICANT: Santasiani, Sima
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

TITLE OF INVENTION: hydroxycinnamic Acid
FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/138, 970A
CURRENT FILING DATE: 2002-07-23
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 22

LENGTH: 716
TYPE: PRT
ORGANISM: Rhodotorula glutinis
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (502) : (502)
OTHER INFORMATION: X=Gly, Ala, Ser, Thr, Pro
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540) : (540)
OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540) : (540)
OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-138-970A-22

Query Match 89.6%; Score 3014; DB 14; Length 716;
Best Local Similarity 83.9%; Pred. No. 3,7e-297;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

```
QY 1 MAPSLDSTATSYXANGXNKGXHAAXXASAXXXXXXAGSXLPTTXXQTOLDIYEXKLAD 60
DB 1 MAPSLDSTSHSANGVASKQAVNGAS-----TNLVAAGSHLPTTQYQYQYDIYEXKLAA 54
QY 61 PXTDDXHELDGSLTLGQVGAARKGRVYXDSDEIRXKIDKSEVEFLRQLSMSYGV 120
DB 55 P-TDSTLELDGSLTLGQVGAARKGRVYXDSDEIRXKIDKSEVEFLRQLSMSYGV 113
QY 121 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSXDFLGLGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSXDFLGLGLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSHAVRLVLEALTNFLNHGITPVLKGTISASGDLSPLSYTAATGHPDSKY 240
DB 174 NSLTRGSHAVRLVLEALTNFLNHGITPVLKGTISASGDLSPLSYTAATGHPDSKY 233
QY 241 HVYHGEKXIKMAREBALFGLPEVVLGPKGGLGVNGTAVASAMATLALDHAMLSLS 300
DB 234 HVYHGEKXIKLAREBALFGLPEVVLGPKGGLGVNGTAVASAMATLALDHAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDYTRPHPTQIEVARNIRTLGSGXFAYHHEEYVKD 360
DB 294 QSLTAMTYEAMVGHAGSFHPLHDYTRPHPTQIEVAGNIRTLGSGRFAYHHEEYVKD 353
QY 361 DEGLRQDRYP-RTSPQWLGPLVSDMIAHAVLSLEAGOSTTNDPLLDVENKXTHHGGNF 420
DB 354 DEGLRQDRYP-RTSPQWLGPLVSDMIAHAVLSLEAGOSTTNDPLLDVENKXTHHGGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 480
DB 414 QAAAVANTMEKTRILAOIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 473
QY 481 AAAYTSELGHLANPYTHVOPAEWNGQAVNSIALISARRTAEANDVLSLLATLHYCVLQ 540
DB 474 AAAYTSELGHLANPYTHVOPAEWNGQAVNSIALISARRTESNDVLSLLATLHYCVLQ 533
QY 541 AYDLRAMEFEFKQFPPXXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTN SYDL 600
DB 534 AYDLRAMEFEFKQFPPXXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTN SYDL 593
QY 601 EPRMHDASFATGTVEXLSXXXXXVSLAAVNAWKVAXAEKASLTRYVXXFXMXPSS 660
DB 594 VERMHDASFATGTVEXLS-----TSLSLAAVNAWKVAXAEKASLTRYVXXFXMXPSS 650
QY 661 SSPALXYLSPTRRVLYSFVREELGVKARGDVFGLGQEVTTIGSVNSRIYEAIKSGRINXV 720
DB 651 SSPALXYLSPTRRVLYSFVREELGVKARGDVFGLGQEVTTIGSVNSRIYEAIKSGRINXV 710
```

Db 651 SSPALSTYLSPTQILIAFAEEIGVAKRGDVFLLGKQZVTIGSNVSKIEAIKSGRINNV 710

OY 721 LVRKMLA 726

Db 711 LVRKMLA 716

Search completed: September 9, 2004, 10:24:27
Job time : 96.3859 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:50:23 / Search time 16.5685 Seconds
(without alignments)
2281.620 Million cell updates/sec

Title: US-09-939-408A-21

Perfect score: 3365
Sequence: 1 MAPSLDLSITSGXANGXNGX.....RIYFAIKSGIKVLYVMA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3021	89.8	1 P15544	rhodospirid
2	2922.5	86.8	1 P10248	rhodocorula
3	1294.5	38.5	1 O93967	amanita mus
4	1205	35.8	1 P35510	arabidopsis
5	965	28.7	1 P45729	petroselinu
6	958.5	28.5	1 P53777	pinus taeda
7	957	28.4	1 P27990	medicago sa
8	954	28.3	1 P17171	oryza sativ
9	952.5	28.3	1 P45734	trifolium s
10	949	28.2	1 P45734	trifolium s
11	947	28.1	1 P45734	trifolium s
12	946.5	28.1	1 P45734	trifolium s
13	946	28.1	1 P45734	trifolium s
14	946	28.1	1 P45734	trifolium s
15	943.5	28.0	1 P45734	trifolium s
16	943	28.0	1 P45734	trifolium s
17	940	27.9	1 P45734	trifolium s
18	939.5	27.9	1 P45734	trifolium s
19	938	27.9	1 P45734	trifolium s
20	933.5	27.7	1 P45734	trifolium s
21	932	27.7	1 P45734	trifolium s
22	931.5	27.7	1 P45734	trifolium s
23	930.5	27.7	1 P45734	trifolium s
24	929.5	27.6	1 P45734	trifolium s
25	929	27.6	1 P45734	trifolium s
26	929	27.6	1 P45734	trifolium s
27	928.5	27.6	1 P45734	trifolium s
28	927.5	27.6	1 P45734	trifolium s
29	927	27.5	1 P45734	trifolium s
30	927	27.5	1 P45734	trifolium s
31	926.5	27.5	1 P45734	trifolium s
32	925	27.5	1 P45734	trifolium s
33	924.5	27.5	1 P45734	trifolium s

ALIGNMENTS

RESULT 1

ID	PALY_RHOTO	STANDARD	PRT	716 AA.
AC	P11544			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Phenylalanine ammonia-lyase (EC 4.3.1.5).			
GN	Rhodospiridium toruloides (Yeast) (Rhodotorula gracilis).			
OC	Eukaryota; Fungi; Basidiomycota; Urediniomycetes;			
OC	Microbotryomycetidae; Sporidiobolales; Rhodospiridium.			
OX	NCBI_TaxID=5286;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CBS 14;			
RC	MEDLINE=92227061; PubMed=1773059;			
RA	Rasmussen O.F., Orum H.;			
RT	"Analysis of the gene for phenylalanine ammonia-lyase from			
RT	Rhodospiridium toruloides."			
RL	DNA Seq. 1:207-211(1991).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IFO 0559;			
RC	MEDLINE=88112870; PubMed=2828184;			
RA	Anson U.G., Gilbert H.J., Oram J.D., Minton N.P.;			
RT	"Complete nucleotide sequence of the Rhodospiridium toruloides gene			
RT	coding for phenylalanine ammonia-lyase."			
RL	Gene 58:189-199(1987).			
CC	- FUNCTION: Catalyzes the nonoxidative deamination of L-			
CC	phenylalanine to form trans-cinnamic acid and a free ammonium ion.			
CC	- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).			
CC	- PATHWAY: Phenylpropanoid biosynthesis; first step.			
CC	- SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),			
CC	which is formed autocatalytically by cyclization and dehydration			
CC	of residues Ala-Ser-Gly (By similarity).			
CC	- SIMILARITY: Belongs to the PAL / histidase family.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib.ch).			
CC	-----			
DR	EMBL; X51513; CA35886.1; -			O9emb9 cicer arlet
DR	EMBL; M18261; AAA33883.1; -			O49835 lithospermu
DR	EMBL; X12702; CAA31209.1; -			S31426 solanum tub
DR	PIR; A29607; A29607.			Q42667 citrus limo
DR	PIR; A56628; A56628.			E31425 solanum tub
DR	HSSP; P21310; 1B8F.			P45730 populus tri
DR	InterPro; IPR008948; L-Aspartase-like.			P45725 arabidopsis
DR	InterPro; IPR001106; Phe/His NH3lyase.			P19143 phaseolus v
DR	InterPro; IPR005922; Phe_am_lyase.			Q43052 populus kit
DR				P45731 populus kit
DR				P4166 ipomoea bat
DR				Q43210 triticum ae

DR Pfam; PF00221; PAL; 1.
 DR TIGRfams; TIGR01226; phe am lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KM Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 211 213 5-imidazolinone (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 212 212 DH4 (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 FT CONFLICT 4 37 SLDSISHSFANGVASAKQAVNGASTLVAAGSHL ->
 FT RPTSGSQRRC (IN REF. 2).
 SQ SEQUENCE 716 AA; 76879 MW; 0CIDF6176944B5E6 CRC64;

Query Match 89.8%; Score 3021; DB 1; Length 716;
 Best Local Similarity 84.0%; Pred. No. 9, 8e-227;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDSTATSXANGXXGXAAXXAXXXXXXAXGSLPTTXXTQDLIVEXXLD 60
 DB 1 MAPSLDSTATSXANGXASAKQAVNGAS-----TLVAAGSHLPTTQVTVIVEXMLAA 54
 QY 61 PXTDDXXELDGYSLTLDVVGAAKRGKRVVXDSDEIRAKIDKSYEFRLXQXNSVYGV 120
 DB 55 P-TDSTLELDGYSLTLDVVGAAKRGKRVVXDSDEIRAKIDKSYEFRLXQXNSVYGV 113
 QY 121 TGFSGADTTEDAISLQKALLHQLCGVLPISXDSFKLGRGLNSLPLEVVRGAMTIRV 180
 DB 114 TGFSGADTTEDAISLQKALLHQLCGVLPISXDSFKLGRGLNSLPLEVVRGAMTIRV 173
 QY 181 NSLTRGSAVRLVLENTNFTLNHGITPIVPRGTSSAGDSPLSYTAAATGHPDSKV 240
 DB 174 NSLTRGSAVRLVLENTNFTLNHGITPIVPRGTSSAGDSPLSYTAAATGHPDSKV 233
 QY 241 HVHGEKRTKMAREAIALFGLSEPVVLGREGGLVNGTAVSASMAATLADAMLSLS 300
 DB 234 HVHGEKRTKMAREAIALFGLSEPVVLGREGGLVNGTAVSASMAATLADAMLSLS 293
 QY 301 QALTALTEAMVGHASFPFLHDVTRPHPTQIEVARNRTLLSESSRAVHHEEVYKDD 360
 DB 294 QSLTAMTEAMVGHASFPFLHDVTRPHPTQIEVARNRTLLSESSRAVHHEEVYKDD 353
 QY 361 DEGILRQDRYPLRTSPQWLGPIVSDMIAHAVALSLBAGQSTTDNPLIDVENKXTHGGNF 420
 DB 354 DEGILRQDRYPLRTSPQWLGPIVSDMIAHAVALSLBAGQSTTDNPLIDVENKXTHGGNF 413
 QY 421 QAAVYNTMEKRTFLATLIGKNTFQLTMLNAGMNRGLPGLAEDSLSYHCKGLDIA 480
 DB 414 QAAVYNTMEKRTFLATLIGKNTFQLTMLNAGMNRGLPGLAEDSLSYHCKGLDIA 473
 QY 481 AAAYTSELGLANPVTTHVQPAEMGNOAVNSIALISASRTAEANDVLSILATLKYCLQ 540
 DB 474 AAAYTSELGLANPVTTHVQPAEMGNOAVNSIALISASRTAEANDVLSILATLKYCLQ 533
 QY 541 AVULRAMEPEPKQEPXXXXLXQFEGKXKTXXXXXLXKXKXKXKLEBQTNSTYD 600
 DB 534 AIDLRALIEEFKKQEPALVSLIDHFGSAMGSLRDELVEKXNTAKLEBQTNSTYD 593
 QY 601 EPRMHDASFANATVVEVEXLSXXXXXSLAAVNAKVAAXEALSLTREVXXKXWXPSS 660
 DB 594 VPRMHDASFANATVVEVEXLSXXXXXSLAAVNAKVAAXEALSLTREVXXKXWXPSS 650
 QY 661 SSPALXYLSPTRTVLVSFYREELGVYKARGDVLGKQVTTGNSVSRITYEAKSGRLNV 720
 DB 651 SSPALXYLSPTRTVLVSFYREELGVYKARGDVLGKQVTTGNSVSRITYEAKSGRLNV 710
 QY 721 LVKMLA 726
 DB 711 LVKMLA 716

RESULT 2
 ID PALY RHORB
 AC P10256; STANDARD; PRT; 713 AA.

DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Rhodotorula rubra (Yeast) (Rhodotorula mucilaginosa).
 OC Eukaryota; Fungi; Basidiomycota; Uredinomycetes;
 OC Microbotrymycetidae; Sporidiobolales; mitospotic Sporidiobolales;
 OC Rhodotorula.
 OX NCBI_TaxId=5537;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL Y-15597;
 RX MEDLINE=89083583; PubMed=3205750;
 RA "Filipula D., Strausberg R.L., Vastle C.A., Sykes A., Levy A.;
 RT "CDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes
 RT reveal genomic intron deletions.";
 RL Nucleic Acids Res. 16:11382-11382(1988).
 CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-
 CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: X13094; CAA31486.1; -.
 CC PIR: S01999; S01999.
 CC HSSP: P21310; 188F.
 CC InterPro: IPR008948; L-Asparaginase-like.
 CC InterPro: IPR001106; Phe/His NH3 lyase.
 CC InterPro: IPR005922; Phe_am_lyase.
 CC Pfam: PF00221; PAL; 1.
 CC TIGRfams: TIGR01226; phe-am lyase; 1.
 CC PROSITE: PS00488; PAL_HISTIDASE; 1.
 CC Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 217 219 5-imidazolinone (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 218 218 DH4 (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 713 AA; 76000 MW; 6EB8317CB037DB8 CRC64;

Query Match 86.8%; Score 2922.5; DB 1; Length 713;
 Best Local Similarity 82.0%; Pred. No. 4, 3e-219;
 Matches 595; Conservative 33; Mismatches 85; Indels 13; Gaps 5;

QY 1 MAPSLDSTATSXANGXXGXAAXXAXXXXXXAXGSLPTTXXTQDLIVEXXLD 60
 DB 1 MAPSLDSTATSXANGXASAKQAVNGAS-----TLVAAGSHLPTTQVTVIVEXMLAA 54
 QY 61 PXTDDXXELDGYSLTLDVVGAAKRGKRVVXDSDEIRAKIDKSYEFRLXQXNSVYGV 120
 DB 60 AGATDQIKLQDGYTLTLDVVGAAKRGKRVVXDSDEIRAKIDKSYEFRLXQXNSVYGV 119
 QY 121 TGFSGADTTEDAISLQKALLHQLCGVLPISXDSFKLGRGLNSLPLEVVRGAMTIRV 180


```

Db      120 TGGGAGDADRTEDDAISLQKALLBHOGLCVLPMSMGCPALGGLGSLNSLPLEVRGAMTIRV 179
QY      181 NSITRGHSARLVVLVLPALTNPLNHGITPIVPIRGITISAGSLSPSYIAAITHPDSKV 240
Db      180 NSITRGHSARLVVLVLPALTNPLNHGITPIVPIRGITISAGSLSPSYIAAITHPDSKV 239
QY      241 HXHXEXEKIMKAREALFGLEPVVLGPKGEGLVNCTAVASAMATLADAMLSLS 300
Db      240 HV--DG--KIMAGQERIALKGLQPVVLGPKGEGLVNCTAVASAMATLADAMLSLS 295
QY      301 QALTALTEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHHEEVKXD 360
Db      296 QALTALTEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHHEEVKXD 355
QY      361 DSGILRODRYPLRTSPQWLGPLVSMIHAAVLSTLGAQSTTNDPLIDVENKXTHHGNF 420
Db      356 DSGILRODRYPLRTSPQWLGPLVSMIHAAVLSTLGAQSTTNDPLIDVENKXTHHGNF 415
QY      421 QASAVXNTEKTRIALALIGKLNFTQLTENLNAKNGRLPSCLAEDPSLSYHCKGLDIA 480
Db      416 MASSVGNTEKTRIALALIGKLNFTQLTENLNAKNGRLPSCLAEDPSLSYHCKGLDIA 475
QY      481 AAATYSELGHANPVTVTHVQPAEMGNQAVNSIALISARPLAENDVLSLLATLYCYLQ 540
Db      476 AAATYSELGHANPVTVTHVQPAEMGNQAVNSIALISARPLAENDVLSLLATLYCYLQ 535
QY      541 AVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYDL 600
Db      536 AVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYDL 590
QY      601 EPRHDAFSXATGVVEXLSXXXXXVSLAANAMKVAARKAISLTXVXXFWXAPSS 660
Db      591 EGRWHDTFVAVGAYVEALAG--QEVSLASLNAKVAQAEKALALTRVDSFWAPSS 647
QY      661 SSGALXYLSPPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRITYEAIKGRINXV 720
Db      648 SSGALXYLSPPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRITYEAIKGRINXV 707
QY      721 LVEMLA 726
Db      708 LVEMLA 713

RESULT 3
ID      PALY ANAMU          STANDARD;          PRT;          740 AA.
AC      093967;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DE      Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN      PAL.
OS      Amanita muscaria (Fly agaric).
OC      Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC      Agaricales; Amanitaceae; Amanita.
OX      NCBI_TaxID=41956;
RN      [1]
RS      SEQUENCE FROM N.A.
RP      Nehls U., Mikolajewski S., Ecke M., Hamp R.;
RT      "Sugar and nitrogen-dependent regulation of an Amanita muscaria
RT      phenylalanine ammonium lyase gene.";
RL      Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
CC      - FUNCTION: Catalyzes the nonoxidative deamination of L-
CC      phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC      - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC      - PATHWAY: Phenylpropanoid biosynthesis; first step.
CC      - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      - PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC      which is formed autocatalytically by cyclization and dehydration
CC      of residues Ser-Ser-Gly (By similarity).
CC      - SIMILARITY: Belongs to the PAL / histidase family.

```

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC      or send an email to license@ebi.ac.uk).
CC      -----
CC      EMBL: A010143; CA009013.1; -.
DR      HSSP: P21310; 1B8F.
DR      InterPro: IPR008948; L-Asparagine-like.
DR      InterPro: IPR001106; Phe/His NH3lyase.
DR      InterPro: IPR005922; Phe_am_1lyase.
DR      Pfam: PF00221; PAL; 1.
DR      TIGRPFAM: TIGR01226; phe_am_1lyase; 1.
DR      PROSITE: PS00488; PAL_HISTIDASE; 1.
KW      Lyase; Phenylpropanoid metabolism.
FT      CROSSLINK 224 226 5-imidazolinone (Ser-Gly)
FT      MOD_RES 225 225 DHA (2,3-DIHYDROQUANTANINE)
FT      FT Best Local Similarity 42.6%; Pred. No. 1,1e-92;
SQ      SEQUENCE 740 AA; 80156 MW; B055CF3D8B7BCEB CRC64;
Query Match 38.5%; Score 1294.5; DB 1; Length 740;
Best Local Similarity 42.6%; Pred. No. 1,1e-92;
Matches 288; Conservative 123; Mismatches 240; Indels 25; Gaps 9;
QY      68 ELDGYSLTLGVDVGARGRGVYVVDSEIRKXIDKSYEFRLQXN--SYGVGTTGFGG 125
Db      69 KVDGQTLTAAVAAARNAAYVLDSEPLVKEVRKSLAIAMKVSFGASYGISTGFGG 128
QY      126 SADRTEDDAISLQKALLBHOGLCVLPSTX--DSFXLGRGLNSLPLEVRGAMTIRVNS 182
Db      129 SADRTDXKMLLGFALLOHVGILPTSTPEPLDVLPLQDANNSTMPAMIRGAILIRNNS 188
QY      183 LTRGHSARLVVLVLPALTNPLNHGITPIVPIRGITISAGSLSPSYIAAITHGDSKYNH- 241
Db      189 LTRGHSARLVVLVLPALTNPLNHGITPIVPIRGITISAGSLSPSYIAAITHGDSKYNH- 248
QY      242 -VXHXEXEKIMKAREALFGLEPVVLGPKGEGLVNCTAVASAMATLADAMLSLS 300
Db      249 GPKSGIRQIGSSKQVLAHNIPEPLSKPEPLGINOTASAVVLAALNEAHLHVLTA 308
QY      301 QALTALTEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHHEEVKXD 360
Db      309 QVTAMGTEALIGTASHAPFIHATARPQOVCAENINMLDGSKLAQLEHEHREVLSD 368
QY      361 DSGILRODRYPLRTSPQWLGPLVSMIHAAVLSTLGAQ--STTNDPLIDVENKXTHHGN 419
Db      369 DXYTLRODRYPLRTSPQWLGPLVSMIHAAVLSTLGAQ--STTNDPLIDVENKXTHHGN 428
QY      420 FQASAVXNTEKTRIALALIGKLNFTQLTENLNAKNGRLPSCLAEDPSLSYHCKGLDI 479
Db      429 FQAMAVTNAMKTRIALALHVGKLFQSTELVNPAMRGPLPSSVAAIDPSPINAHAKGLDI 488
QY      480 AAATYSELGHANPVTVTHVQPAEMGNQAVNSIALISARPLAENDVLSLLATLYCYL 539
Db      489 ATPAAVAAE--ATGPFTHIQSAEKNQAVNSIALISARPLAENDVLSLLATLYCYL 544
QY      540 QAVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYD 599
Db      545 QAVDLRALQREFLPGLDIIIRELRSFGFLS--SEQNEKIQCVLTAFAEHDHDKTTMD 603
QY      600 LEPHMDAFSAATGTVVEXLSXXXXXVS-----LAANAMKVAARKAISLTXVEXKF 654
Db      604 NTDFTMTAATSSVLTLOFTDSCGAVPSSCDLLSSVSFQSSVATRRSVLMDLKEY 663
QY      655 WAPSSSPALXYLSPPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRITYEAIKGR 708
Db      664 TRGDGEPPLPASQYIG-KTRPYIQFIRTTIGVRKSGSEVYNKFNGLGVEDVTIQNSIRI 722
QY      709 YEAIKGRINXVLMV 724

```

```

Db      723 YESIRGDMOSIIVSL 738

RESULT 4
PALY_USTWA STANDARD; PRT; 724 AA.
ID_PALY_USTWA
AC 096V77;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PALI.
OS Ustilago maydis (Smut fungus).
CC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
CC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim S.-H., Macdonald K., Virmant D., Wake K., Kronstad J.W.,
RA Ellis B.E.;
RT "Cloning and disruption of a phenylalanine ammonia-lyase gene from
RT Ustilago maydis.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF306551; AL09388.1;
DR InterPro; IPR008948; L-Asparatase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfams; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism.
FT CROSSLINK 205
FT 5-imidazolone (Ala-Gly)
FT (By similarity).
FT MOD_RES 206 206 DHA (2,3-DIDEHYDROALANINE)
FT (By similarity).
FT (By similarity).
SQ SEQUENCE 724 AA; 79277 MW; 02A73C97C34C01AB CRC64;

Query Match 35.8%; Score 1205; DB 1; Length 724;
Best Local Similarity 41.6%; Pred. No. 9,1e-86;
Matches 287; Conservative 114; Mismatches 245; Indels 44; Gaps 15;

QY 69 LDGVSLTLDGVGAARKGRXVYXDSDEIRXKIDSVKRLXQV--XNSVYGVTTGGGS 126
DB 41 IDGNLKLKIGLVASAYGVHTPRPSAETRKRIDSVQSLAKLDGSGSITGINTGGGS 100
QY 127 ADRTEDALSQKALLHQLCGLVPT-----SXDSFXLG-RGLENSP--PLEVVRG 174
DB 101 ADSRTANTRALQALHMOOCGLVPTSTFTGESSAPFALPLTDTSSSLMEAVVRG 160
QY 175 AMTIRVNSLTGRGSANRVLVLELNT-FLNHGITIVPLRGITSSGSLSLSTIAAIT 233
DB 161 AIVVRSLSSLMRGSGVMEVLDMDQLFLQNNVTVPVVRSSISASGSLSPSLVAAGLA 220
QY 234 GHDPDSKVHVHEGEXKI-MKAREALFLGLEPVVLGPEKGLVNGTAVASMAATLALHD 292

```

```

Db      221 GORGIYCFVTDGRGRVKTADACRMAKITPVQYEEPEVGLNGTAFSAVGLATYE 280
QY 293 AMLSLISQALPTALTEAVMGHAGSFHPLHDVTRPHPTQIEVANRTITLGGXFVH- 351
DB 281 AENIASLTQLTAVAVEALKGTIDASFAFPIHEIARPHGQIKSAKFIKRAHLSGRLEHL 340
QY 352 -HEEYKVKDDEGITRORRYPLRTSPQWLGSLVDMTHAHAVLSLEAGSTTNPLDV- 409
DB 341 ENEKVLFSEBDGTLRQRYTLQRTASQWVGGLDIENAKSVDFEI--NSTDDPMIDPY 399
QY 410 -ENKXTHHGNGFQASAVVNTWEKTRTALALIGKINFTQTEMLNAGNRGIPSCIA-ED 467
DB 400 DDDGRHHGNGFQAMAMNNAVEKIRLALCAQKMTFQGMTELVNPNRRGLPAILSTPD 459
QY 468 PSLSTYHCKGLDIAAAAYSEIHLANPYTHVQPAENGQAVNSLALISARPAEANDVL 527
DB 460 LSLNFKAGINIALASVTSSELMFLGNPVTHTVQSAEVANQAFNSLALISGRQTLQATECL 519
QY 528 SLLLAHTLYCYQAAYDLAMEPEFKQEPXXXXLLKHQFGKXTXXXXXELXXKXKX 587
DB 520 SWIQWSLYLLQALDIALDQYKABQLPTILASLSHRS-EMMDETKQGTIAQVLS 578
QY 588 LKRLLEQINSYDLBPRMHDASFATGTVEXLSS--XXXXXVSLAANAKVAXAEKALS 645
DB 579 MKKRLDETSKDLRDLVETVQDASSVLYRFSELPGGCGADPLRNIVKRWATGVADTEK 638
QY 646 LTRKYRXXFWKXAPSSSPALKYLSPRTRVYXSFVREELGYAKRGDVFLEKQEV----- 699
DB 639 ITRQVTEIEDNPVYCHAS--HLGKTRGAEEFVKTLGVPMH-----GKSNLNEPKGE 690
QY 700 -----TIGSNVRIYEAIKSGRINXVLVKM 724
DB 691 PEQWNTGGYVSVIYASIRDEGLVNMKSEL 720

RESULT 5
PALI_ARATH STANDARD; PRT; 725 AA.
ID_PALI_ARATH
AC P35510; G92QD6;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase I (EC 4.3.1.5).
GN PALI OR AT2G37040 OR TUB8.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Arabidopsi: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=95195160; Pubmed=788622;
RA Wanner L.A., Li G., Ware D., Somsich I.E., Davis K.R.;
RT "The phenylalanine ammonia-lyase gene family in Arabidopsis
RT thaliana.";
RL Plant Mol. Biol. 27:327-338 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; Pubmed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.W., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.-J., Ronning C.M., Xoo H.L.,
RA Moffett K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Frazer C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768 (1999).
RN [3]

```

```

RP SEQUENCE OF 1-240 FROM N.A.
RC STRAIN=cy. Columbia;
RA MEDLINE=93005677; PubMed=2152131;
RX Ohl S., Hegrick S.A., Chory J., Lamb C.J.,
RT "Functional properties of a phenylalanine ammonia-lyase promoter from
RL Arabidopsis.",
RT Plant Cell 2:837-848(1990).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J36177; AAC18870.1; -.
DR EMBL; AC006922; AAM15324.1; -.
DR EMBL; X62747; CAA44609.1; -.
DR PIR; S52990; S52990.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRPFAM; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KM Lyase; Phenylpropanoid metabolism; Multiligene family.
FT CROSSLNK 211 213 5-imidazoligene (Ala-Gly)
FT MOD_RES 212 212 DHA (2,3-DIDEHYDROALANINE)
FT CONFICT 329 329 V -> I (IN REF. 2).
FT CONFICT 426 426 R -> A (IN REF. 2).
FT CONFICT 612 612 V -> I (IN REF. 2).
SQ SEQUENCE 725 AA; 78782 MW; 5D65F10D50A2946 CQC64;

Query Match 28.7%; Score 965; DB 1; Length 725;
Best Local Similarity 35.5%; Pred. No. 3.9e-67;
Matches 258; Conservative 119; Mismatches 258; Indels 92; Gaps 21;

QY 40 GSKLPTTXXTQDLYEXKLADPXTDDXXELDGYSLTLDGVGAARKGRKXRVXSDSIRX 99
DB 39 GAAEQMKSHLDEVKRVVAE-FRRPVNLTGGETLTTCQVAIASTIGNSVVELSETPARA 97
QY 100 KIDKSVFPRXQLOXNSV-YGVTFGRGSDRTEDATSLXALIEHQLCVLPTSDSF 157
DB 98 GVAASDWVMSKMGTDGYGTTTGFATSHRTKNGVALQKELIRPLNGITPSTKET- 156
QY 158 XIGRGLNSLPLEVVGANTIRVNSILTRGSAVRLVLVLTNPLNIGTPIVPLRGTS 217
DB 157 -----SHTLPSATFAALVLRINTLQFGSIRFEIEIATISFLINNITPLSLRGIT 210
QY 218 ASGDLSPLSYIAAATGHDSDRVVYHHEKXIKVAKAIALFGLEP--VVLGREGIGL 275
DB 211 ASDDLPLSYIAGLILGRPNKA-TGPNQ--EALTAENFLKAGISSGFFPLQKREGIAL 267
QY 276 VNGTAASAMATLALHDAHLSLSQALTALTVEAMVGHAGSFPHLDVVRPHPTQIEV 335
DB 268 VNGTAAGSGASGVLFETNVLSVLAELISAVPAEWSGKP-EFTDHLTRHKHPCQIEA 326
QY 336 ANRIRTLRGSXF-----AVHHEEVKVKDDBGILQDRYPLRTSPQMLGRLVSDMIRHNA 391

```

```

DB 327 AAVMEHILDGSSYNKLAQKHENDPLOKP-----KODRVALRTSPQMLGFOIE-VIRYA 379
QY 392 VLSLEAG-QSTTDNPLIDVENKTHGNGFQASAVXNMTKELTALALIGKLNFTOLTEN 450
DB 380 TKSIERELNSVNDPFLIDVGRNKAHNGNFQGPICGVMONTLALRALKLMFAQPSSEL 439
QY 451 LNAGNRCLPSC-L-PAEDPSISYCHCKGLDIAAAYSELGHLANPPTVTHQPMENQAV 509
DB 440 VNDPYNMGLPNNLTASRNPISLDYFPGAEIAMASYCELOLYLANPVTSHVQSAEQHNQDV 499
QY 510 NSLILISARCTAEANDVLSLLATHTLYCVLQAVDLRAMEFEFKQFEXXXLXQHFGX 569
DB 500 NSLGLISRKSEAVDILKMTSTFLVAILQAVDLRLLENLAQ----- 543
QY 570 XXXTXXXXXXELXKRV-----XXKIXKRL-EQNSYDLERPMADPSKAT 612
DB 544 --TVANTVSQVAKVLLTGVNGELHPSRFCEKDLTKVDDEQVYVADDP-----CSAT 595
QY 613 GTVVEKXLSXXXXXVSLAANA-----WKVXAKKALS--LTRVYRXXFWXAPS 659
DB 596 YPLQKLNQ--VVDHALVNGSEKNAVTSIFRKGAFEEELKAVLPKVEAARAAYDN 652
QY 660 SSSPALXLYSP-RTRVLYSFVREELGVKARGDVLGKQEVTTGNSVRYEAIKSGRN 718
DB 653 GTSALPNRIKRCRQVPLRVFVREELGT-----ELTGEKVTSPGEEDKVFALICEGKI 707
QY 719 XVLVNM 725
DB 708 DEWMECL 714

RESULT 6
PAL3_PETCR STANDARD; PRT; 718 AA.
ID PAL3_PETCR
AC P45729;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 3 (EC 4.3.1.5).
GN PAL3;
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; asterids;
OC campanulids; Apiales; Apiaceae; Apioidae; aploid superclade;
OC Apium clade; Petroselinum.
OX NCBI_Taxid=4043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95010141; PubMed=7925471;
RA Appert C., Logemann E., Hahlbrock K., Schmid J., Amrhein N.;
RT "Structural and catalytic properties of the four phenylalanine
RT ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL Eur. J. Biochem. 225:491-499(1994).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81159; CAA57057.1; -.

```

PALY	PINTA	STANDARD	PRT	754 AA
ID	PALY_PINTA		PRT	754 AA
AC	P52777			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Phenylalanine ammonia-lyase (EC 4.3.1.5).			
GN	PAU.			
OS	Pinus taeda (loblolly pine).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.			
OX	NCBI_Taxid=3352;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Needle;			
RA	Zhang X.H., Chiang V.L.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the			
CC	first reaction in the biosynthesis from L-phenylalanine of a wide			
CC	variety of natural products based on the phenylpropane skeleton.			
CC	-1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).			
CC	-1- PATHWAY: Phenylpropanoid biosynthesis; first step.			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (probable).			
CC	-1- PIM: Contains an active site 4-methylidene-imidazole-5-one (MIO),			
CC	which is formed autocatalytically by cyclization and dehydration			
CC	of residues Ala-Ser-Gly (by similarity).			
CC	-1- SIMILARITY: Belongs to the PAL / histidase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U39792; AA84889.1; -			
DR	PIR; T09777; T09777.			
DR	HSSP; P21310; 1B8F.			
DR	InterPro; IPR008948; I-Aspartase-like.			
DR	InterPro; IPR001106; Phe/His_NH3lyase.			
DR	InterPro; IPR005922; Phe_am_lyase.			
DR	Pfam; PF00221; PAL; 1.			
DR	TIGRFAMs; TIGR01226; phe_am_lyase; 1.			
DR	PROSITE; PS00488; PAL_HISTIDASE; 1.			
FT	Lysase; Phenylpropanoid metabolism.			
FT	CROSSLINK 206 5-Iminidazolone (Ala-Gly)			
FT	MOD RES 207 207 (By similarity).			
FT	DHA (2,3-DIHYDROALANINE)			
FT	(BY SIMILARITY).			
FT	SEQUENCE 754 AA; 82600 MW; 59E97466AAC398 CRC64;			
SO				
Query Match	28.4%; Score 957; DB 1; Length 754;			
Best Local Similarity	36.9%; Prod No. 1.7e-66;			
Matches 252;	Conservative 109; Mismatches 258; Indels 64; Gaps 18;			
QY	LDGSLVLTGVDVGAARKGRVYXVDSDEIRKXIDNSVEFLRXQLX--NSVYGVTTGGGS 126			
Db	64 IEGSLTLTSDVAAVAARRSQVKVLLDAAAKSRVEESNMVLLQMTKGTDTGVTTGGAT 123			
QY	127 AQTFRDEALISLQKLLLEHLOGVLPTNSXSFKLGGLENSLPLEVYVRAMTIRNSLTIRG 186			
Db	124 SHRRINQGAELQKELIRFLNAGV-----LGKPEVNLSEDTTRAMLVRRNTLLQG 174			
QY	187 HSAVELVVLALTLFNLHGHTPIVPLRGTTISAGDLSPISTYAAITGHPDSKVYVHNG 246			
Db	175 YGSGVMDLLEIVETKLLNVAWLTPLRLRGTTISAGDVLPSYLAGLTGRPSVVR--SRDG 223			
QY	247 XKTKXNAEBAALFGLF--PVYLPGRKGLGVNNGTAVSASMATLALDHAATLSLSQALT 305			
Db	234 IF--NSGAELAKVYGLKPEFLQPKKGLAIWGTISVGAALSIYCFDPAIVALLSEVISA 291			
QY	306 LIVEAMVGHASFPFLHDTVRPHQIIVANIRITLLEGSAFAVH-----HEEEVKYKD 361			

```

Db      292 MCEVWNGPEFTDYLTHKL-KHPGQWEAALMEYVLVDGSSYKKAUKHEMNPLOXP- 349
Qy      362 EGILRODRYPLRTSPQWLGPLYSDWIMAHAVLSLSEAGOSTDNPDLIVENKXTHHGNFQ 421
Db      350 ----KQDRGLRTSPQWLGPLYSDWIMAHAVLSLSEAGOSTDNPDLIVENKXTHHGNFQ 404
Qy      422 ASAVXNMEKTRTLALIGKLNFTLTMLNAGNMGIPCLTA-AEDPSISYCKGIDIA 480
Db      405 GTPIGSDMNLNLISALGKLMFAQFSLVNDYVNGJLPSYLSGGFPNPSLDYGLKGAIEA 464
Qy      481 AAAYTSELGLHNPATYTHVQAPMEMNQAVNSIALISARTEANDVLSLTLATHYCVLQ 540
Db      465 MASTSELHLYNPSTSHVQSAEQHNQDVNSIGLVSAKSAEALDILKMLSTYLTLCQ 524
Qy      541 AVDLRAEFEPKQFPKXXKXLLXQHPGXXXVXXXXXXELXX-VKXKXKREQTN--- 566
Db      525 AVDLRLHE----ENMLATVQIVSQVAKKTLSTGLNGSLDLPGRCEKDLQVVDNEHVPF 580
Qy      597 -----SYDLEPMMHD-----ARSAATGTVEKLSXXXXXXVSLAANAMKVAAXEKA 643
Db      581 YDDPCNMSYPLTQKLRNLVHFAFKABEGEDPNTSIFNKI PVPFAELK---AQLEPQ 636
Qy      644 ISLTRVXXXFXWAPSSSPALXVLP-RTRVLSFVAREELGVARRQDVFELGQEVTTIG 702
Db      637 VGLARE-----SYDKGTSPLPDRIGECRSYPLVEFVANGQGTK-----LLSGTRTISPG 685
Qy      703 SNVSRLEYAKSGRINXVLYKVL 725
Db      686 EVIEVYDAISDQKVIPLFKCL 708

RESULT 8
PALY_MEDSA
ID      PALY_MEDSA      STANDARD;      PRT;      725 AA.
AC      P27990;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      phenylalanine ammonia-lyase (EC 4.3.1.5).
OS      Medicago sativa (Alfalfa).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosid1; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.
OX      NCBI_TaxID=3879;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Apollo;
RX      MEDLINE=91355933; PubMed=1715786;
RA      Gower G., Patra N.L., Dixon R.A.;
RT      "Stress responses in alfalfa (Medicago sativa L.) 12. Sequence
RT      analysis of phenylalanine ammonia-lyase (PAL) cDNA clones and
RT      appearance of PAL transcripts in elicitor-treated cell cultures and
RT      developing plants.";
RL      Plant Mol. Biol. 17:415-429(1991).
CC      -|- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC      first reaction in the biosynthesis from L-phenylalanine of a wide
CC      variety of natural products based on the phenylpropane skeleton.
CC      -|- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC      -|- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC      -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      -|- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC      which is formed autocatalytically by cyclization and dehydration
CC      of residues Ala-Ser-Gly (by similarity).
CC      -|- SIMILARITY: Belongs to the PAL / histidase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

```

CC      EMBL: X58180; CAA41169.1; -.
DR      PIR; S17444; S17444.
DR      HSSP; P21310; 1B8F.
DR      InterPro; IPR008948; L-Asparase-like.
DR      InterPro; IPR001106; Phe/His NH3lyase.
DR      InterPro; IPR005922; Phe_am_lyase.
DR      Pfam; PF00221; PAL; 1.
DR      TIGRfam; TIGR01226; phe_am_lyase; 1.
DR      PROSITE; PS00488; PAL_HISTIDASE; 1.
KW      Lyase; Phenylpropanoid metabolism.
FT      CROSSLINK 211 213 5-imidazoleinone (Ala-Gly)
FT      MOD_RES 212 212 DHA (2,3-DIDEHYDROALANINE)
FT      MOD_RES 212 212 (BY SIMILARITY).
SQ      SEQUENCE 725 AA; 78665 MW; 19906A2575F64D7D CRC64;
Query Match 28.4%; Score 954; DB 1; Length 725;
Best Local Similarity 36.2%; Pred. No. 2,86-66;
Matches 262; Conservative 106; Mismatches 252; Indels 104; Gaps 21;
Qy      49 TOLDIVEXKXADPXTDDXKBLDGYSLTLDGVGAKRGKRYKXVDSPEIRKXKIDKSYEFL 108
Db      48 SHLDVEMKMAVE-YKPVVRIGGETLTIQVAAIAADHGVQVLDSSARQDVASSEWV 106
Qy      109 RXQLKNSV--YGVTTGFGGSADTRTEDAISQKALLHEQLGVLPSTXDSFXLGRGLENS 166
Db      107 MESWKKGTDSYGVTTGFGATSHSRKQGGALQKELIFLNAIGTNGTES-----NHT 159
Qy      167 LPLEVVRGAMTIRVNSLIRGSAVRLVVLALTNFLNHGITPIVPLRGTTISAGDLSPLS 226
Db      160 LPKTRRAAMLVIRINTLLQSGIDFELLEAITKPLNKVTVPCLPLRGTTISAGDLSPLS 219
Qy      227 YTAALITGHPSKXVHHEGKEXIMXAREALALFGL--EPVVLGPKRGKLGIVNGTAVSAS 284
Db      220 YTAGLTLGRPNK--AHGSGEVLNAKEAPNLAINAPFELQPKGGLAVNGTAVAGSG 276
Qy      285 NATLALHDAHMLSLSQLTALTVEAWGHAASFPHLDVTRPHPTQIEVARNITLLE 344
Db      277 LASTYLFEENILAVLSEVLSAIFAEVWQGR-EFTDHLTHLKHHPQIEAAALNEHILD 335
Qy      345 GSKFAVH---HEEVKXKYKDEGLIRQDRYPLRTSPQWLGPLYSDWIMAHAVLSLEAG-Q 399
Db      336 GSGYVAAKKLHEIDPLQKP-----KQDRVALRTSPQWLGPLYE--VIRFTKSIIEEIN 388
Qy      400 STTDNPDLIDVENKXTHHGNFQASAVXNMEKTRTLALIGKLNFTLTMLNAGNMGRL 459
Db      389 SYNDNPDLIDVSRNKALHGNFQGTPIGVSMNTRLALASIKLMFAQFSELVNDPYNGL 448
Qy      460 PSQL-AAEDPSISYCKGIDIAAAAYTSELGLHNPATYTHVQAPMEMNQAVNSIALISAR 518
Db      449 PENLSASRNPISLDYGFKAETAMASYGSELQYLANPVTTHVQSAEQHNQDVNSLGLISAR 508
Qy      519 PRAEANDVLSLILAHLYCYQAVDLRAEFEPKQFPKXXKXLLXQHPGXXXVXXXXXX 578
Db      509 KNEAEILEIQLMSSFFLLALQQAIDRLHLENLN----- 543
Qy      579 ELKXKXKXKXKRLQTSYDLEP-----RWH-----DAFSKATGTVEKLS 620
Db      544 SYKNTVSVQAKKTLTMGVNGNELHPSRFCEKDLKVVDREHFAVYIDPCSAITYPLSQGLR 603
Qy      621 SKXXXXXVSLAANA-----WKVAXAEKALS--LTKVFXKFXWAPSSSPALXV 667
Db      604 Q---VLVDHALNGSEKKNFNTSIFOKIATEEELKTLPLPEVSA-RTAVESGNPTI-- 657
Qy      668 LBP-----RTRVLSFVAREELGVARRQDVFELGQEVTTISNVSRIYEAISGRINXV 721
Db      658 --PNKINGRSIPLYKVFREELGT-----GLTGBNVISPGECDKLFAMCGKIIDPL 710
Qy      722 VVQL 725
Db      711 LECI 714

```

RESULT 9
 PALL_ORISA STANDARD; PRT; 701 AA.
 ID PALL_ORISA
 AC P14717
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 OS PAL.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzae; Oryza.
 OC NCBI_Taxid=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA MEDLINE=90032676; PubMed=2806257;
 RX Mitani E., Ozeki Y., Matsuo M., Kozuka N., Tanaka Y.;
 RT "Structure and some characterization of the gene for phenylalanine
 ammonia-lyase from rice plants.";
 RL Eur. J. Biochem. 185:19-25(1989).
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 first reaction in the biosynthesis from L-phenylalanine of a wide
 variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 which is formed autocatalytically by cyclization and dehydration
 of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16093; CA34226.1; -.
 DR HSSP: P21310; 1B8F.
 DR Gramene; P14717; -.
 DR InterPro: IPR008948; L-Aspartase-like.
 DR InterPro: IPR001106; Phe/His NH3lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; Pal; 1.
 DR TIGRfam: TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 189 191
 FT MOD_RES 190 190 DHA (2,3-DIDEHYDROALANINE)
 FT (By similarity).
 FT (By similarity).
 FT MOD_RES 190 190 DHA (2,3-DIDEHYDROALANINE)
 FT (By similarity).
 FT (By similarity).
 SQ SEQUENCE 701 AA; 75808 MW; 2BF9ADP0BCE8E4DE CEC64;
 Query Match 28.3%; Score 952.5; DB 1; Length 701;
 Best Local Similarity 35.9%; Pred. No. 3.5e-66;
 Matches 254; Conservative 117; Mismatches 266; Indels 71; Gaps 21;
 QY 40 GSKLPTXXHOLDIVXKXADPXTDDXXHLOGYSITLGDV--VGAARKGRVRYXSDDEI 97
 DB 16 GAATAAAGSHLDEVMGWAQ--FREPLVKTQATIRVQVAAVQAQDAARVAVELEDEA 74
 QY 98 RKXIDKSVFLRXOLAN--SVYGVTTFGSGADRTEDPAISLQVALLHOLCGVLPSTXD 155
 DB 75 RPRVKAASWMLTCAHGGDIDVGTITGGTSHRTKQGPALQVELRLYNAGAFGTGSD 134
 QY 156 SFTLGGLENSLPLEVVRGAMTIRVNSLTRGSAARLVVLAALNPLNHGCTPVPPLRGT 215

Db 135 G-----HTLPSETRVAAMLVRINTLLQGYSGIREILLEATIKLINTGTPCLPRGT 186
 QY 216 ISAGDLSPLSTIAAIIIGHPSKRVHYHEGKXIKMAREALAFGLEP--VLIGREG 273
 Db 187 ITASGDVLPSTIAILLITGRPWAQ--ISPDRK--VDAAEFKLAGEGFTLNPREGL 243
 QY 274 GIVNGTAVASAYATLADHMLSLSQALATLVEAMVAGSFFHLDVTRPHTQI 333
 Db 244 AIVNGTSOSALAAVMDANILAVLSVLSAVCEVMNGPEYTHLTKL--KHPGSI 302
 QY 334 EVARNIRTLBGSXAFAHHEEVEKYKDDGL--RQRYPLRISPOMLGVLSDMTHAA 391
 Db 303 DAAAMEHILAGSSFSMSHAK--KVENMDPLKXQDRYALRTSPQWLGPQI--QVIRA-A 357
 QY 392 VLSLEAG--QSTDNLDIVENKXTHHGGNFPASAVXNMTKRTALALIKLNFOTLEM 450
 Db 358 TKSIREVNSVNDPVIDVHKGKALHGENPFGTIGVSMNARLAINIKLMFAQPSL 417
 QY 451 INAGNNRGLPGLA-AEDPSLSYHCKGLDIAAAVTSSELGHLANPVTTHVQAPAMGNAV 509
 Db 418 VIEFYNNGLTSLAGSRNPSLDYGFKEIIMASYSSELQYLANPITNHVQSAEONQDV 477
 QY 510 NSALISARRTEANDVSLTLATHIVCVQAVDLRAMEEFPKQFPKXXXLLXQHPGX 569
 Db 478 NSGLVSAKTLAEVDILKMTSTYIALCQAVDLRHLHENIYSSVKNCTQYAKKVLTM 537
 QY 570 XXXXXXXXELXXKVKXKXLEQTNSTYDLEP--RWHDAFSXATGVVE-XLSXXXXX 626
 Db 538 NTGDLSSARSEKQKLLAIR--EAVSYADDPESANYPMLQKRLAVLVEHALTSGDRA 596
 QY 627 VSLAIVNAKRYAKXAEKALSLTRVXXKFWAPSSSP-----ALVYLS 670
 Db 597 RGLRLYQDHYRG--APLCAAGDGRPRRRRORTAPYANRIVS 640
 QY 671 RTRVLYSFREELGVKARGDVPF--GKQEVITGSNVRIRIYAKSGRI 717
 Db 641 RSPFYRFRVREELG-----VFLTGERKLSPECEKXVLGISQK 682
 RESULT 10
 PALLY TRISU STANDARD; PRT; 725 AA.
 ID PALLY TRISU
 AC P45734;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PALLY.
 OS Trifolium subterraneum (Subterranean clover).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
 OX NCBI_Taxid=3900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Karriadi; TISSUE=leaf;
 RX MEDLINE=94171061; PubMed=8125321.
 RA Howles P.A., Arioli T., Weirman J.J.;
 RT "Characterization of a phenylalanine ammonia-lyase multigene family
 in Trifolium subterraneum.";
 RL Gene 138:87-92(1994).
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 first reaction in the biosynthesis from L-phenylalanine of a wide
 variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 which is formed autocatalytically by cyclization and dehydration
 of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, M01192; AAA17993.1; -
 DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Asparatase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRPFAM; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 211 213 5-imidazolone (Ala-Gly)
 FT MOD_RES 212 212 (By similarity).
 FT DHX (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 FT SEQUENCE 725 AA; 78998 MW; 70F9925C6622240 CRC64;
 SQ
 Query Match 28.1%; Score 949; DB 1; Length 725;
 Best Local Similarity 37.2%; Pred. No. 6,7e-66;
 Matches 252; Conservative 108; Mismatches 266; Indels 52; Gaps 18;
 QY 69 LOGSVLTLDVVG-AARKGRXVXVDSDEIRKIDKVEFLRXOLXNSV--YGVTTGFG 125
 DB 68 LGGETLTISQVVAIAAHDCATVEL--SSARAGVYASSDWMESMKNKTSYGVTTGFGA 125
 QY 126 SADTRTEDAISIOLKLEHQLGVLPFSXDSFLGRGLNSLPLEVYRGANTIRVNSLTR 185
 DB 126 TSHRRTRKQGGALQKELIRFLNAGIRGNGES-----NHTLPATRAALVRIINTLQ 178
 QY 186 GHSAYVLVLEALTNFNLNGITPIYPLRGTTISAGDLSPLSYIAATIGHDSKRVYHE 245
 DB 179 GSGIRFELLEIAITKLNNNTIPCLPLGTTASGDLPVLSYIGLGRSNSK--AHG 235
 QY 246 GKEIKXAREEALFGL--EPVVLGPKEGELVNGTAVASAMATLADHMLSLSQL 303
 DB 236 PSEEMINAEQAOLGINSFELQPKEGIALVNTGAVGSGLASIVLEANIILVLEVL 295
 QY 304 TALTVAMTGAHSGRFPFLADVTRPHPTQIEVAKIRFLLESGKFAVHNEEVKXVDEG 363
 DB 296 SAIFAEVWQGR-EPFDHLYHKLKHPGQIEAAIMELHLSGAY-VYDAKQLEMPLO 353
 QY 364 LRQDVEPTSPQWGLPVSDIMIAHVLSEAG-OSTDNPILDVKNKXTHGNGFOA 422
 DB 354 KPRQDRIYALRTSPQWGLPFI--VIRFSTKSIEREINSVNDNPILDVGRNALHGSNFG 411
 QY 423 SAYANTMEKTRIALALIGKLNFTQTEMNAGNGLPSCU-AAEDPSLYHCKGLDIAA 481
 DB 412 TPVGSMNDTRLALASIGKLPFAQSELVNDPYNNGLSNASHNPFLDYGFGSEIAM 471
 QY 482 AATTSGLHANPVYTHVQPAEMGNQAVNSLALISARRTAENDVLSLLATHLVCLQA 541
 DB 472 ASYCSLOYLANPVTHVQSAQHNDVNSLGLISSRKTKEAIEILLQMSSTFLIALCQA 531
 QY 542 VDIRAMEFEKQGFXXXXLXKHGXXXXTXXXXXXKXKXKRLQEGNSYDL 601
 DB 532 IDIRHLEENIKSVKNTVQVAKKITITIGVSEGLPSPFCEDLLKAVDR-EHVFSTIDD 590
 QY 602 PRWHDASXATGTVEVXLSXXXXXVSLAANA-----WKVAXAKAAS-LTR 648
 DB 591 P-----CSATYPLAQKLRQ--VIVDHALVNGSEKSNSTISFOKIAFEEELKTLIPK 641
 QY 649 XGVXXVWAPSSSSPALXVLSF-KTPVIXSFVREELGVKARQGVFLGKQVTTGSNSR 707
 DB 642 EVESAPTAENGSTANKINGCRSTPLKVFVEELGT-----SLTGERVISPGEBCDK 696
 QY 708 IYEAISGRINXVLM 725
 DB 697 LFTAMQCGKIDPLKCL 714

RESULT 11
 ID PAL1 PRUAV STANDARD; PRT; 717 AA.
 AC 064963;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase I (EC 4.3.1.5).
 GN PAL1.
 OS Prunus avium (Cherry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid I; Rosales; Rosaceae; Amygdaloideae; Prunus.
 OX NCBI_TaxID:42229;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summit;
 RA Wiersma P.A., Wu Z.;
 RT "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe
 RT Sweet Cherry fruit (Prunus avium).";
 RL (in) Plant Gene Register, p698-184.
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF036948; AAC78457.1; -
 DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Asparatase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRPFAM; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 203 205 5-imidazolone (Ala-Gly)
 FT MOD_RES 204 204 (By similarity).
 FT DHX (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 FT SEQUENCE 717 AA; 77999 MW; B64DF9FADBF0B3 CRC64;
 SQ
 Query Match 28.1%; Score 947; DB 1; Length 717;
 Best Local Similarity 35.8%; Pred. No. 9,5e-66;
 Matches 250; Conservative 112; Mismatches 285; Indels 52; Gaps 17;
 QY 40 GSXLPITXTQTDIYEXKLABPXTDDXXELDQYSVTLGDVYGARKGXVXVXSDERX 99
 DB 31 GVAAEFLKGSNHLDEYKRWVAE-YRKPVVKLGESLTISQVAIAITHDSGVVEIASEARA 89
 QY 100 KIDKSVFELRXOLXNSV--YGVTTGFGSADTRTEDAISIOLKLEHQLCGVLPFSXDSF 157
 DB 90 GYKASSDWMDMSMGTSYGVTTFGATSHRRTQGAALQKELIRFLNAGVFPSTKES- 148
 QY 158 XLGRGLNSLPLEVYRGANTIRVNSLTRGHSANRIVLEALTNFNLNGITPIYVLRGIS 217
 DB 149 -----GHTLPQCATRAALVRIINTLQGYSGIRFELLEIVITKFLNNNTVCLPLRGTT 202

QY 218 ASGDLSTSYIAAATGHPDKVYKHHEGKIMKAREALFGLER--VVLGREGJGL 275
 DB 203 ASGDLVPLSTIAGMLTIRPNSKA-VGPBG--QTLAAEPEFVINGSGFEELQPEGLAL 259
 QY 276 VNGTAVASAMATLALDHANLISLSQALPTLVEAMVAGAGSFHPLDVTREPHPTOLEV 335
 DB 260 VNGTAVAGSGLASVTLFTDITLITALLSELISAFEVNQGKR-EFTDHLTHKHKHNGQLEA 318
 QY 336 ARNRTLLGSGKFAVH---HEEEVAVKDEGLRDPRPLTSPQMLGPLVSDMTHA 391
 DB 319 AALWEHTLDSSSYVAKKAKKLEQDPQKR---KQDRALRTSPQMLGPOLE--VIRIS 371
 QY 392 VLSLEAG-QSTNDPLIDVENKCTHGGNFQASAVNTMEKTRMLALIGKLNFTQTEM 450
 DB 372 TKSLERIDSVNDPLIDVSRNKALHGNFGCPISVSDNRLRLAIGKLMFQPSL 431
 QY 451 LNAQWNGPLSPCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAENQAV 509
 DB 432 VNDPYNNGPLSNLSGGNPSLDYGFKAELIAMAASYCELQFLANPVTNHVQSAEQHNDV 491
 QY 510 NSIALISARRTAAANDVLSLLATLHYCVLQAVDLPAWFEFKQFQXXXXLKHGX 569
 DB 492 NSIGLISSRKTAABVDLKLMSSTFLVALCQALDLHLEENLRNTVKNVQVAKR--T 548
 QY 570 XXTXXXXXXELXXKXXKXKLEQTNYSYDLPPRHDAFSXATGTVEIXSSXXXXVSL 629
 DB 549 LTTGVNGELHPSRFCEKDLKVDREYV---AYIDPCSAITYLMQKLVLEHNT 604
 QY 630 AAVYAMKVAA--EKALSTLTXVXXFKWAPSSSPALXYUSP-----FTRVLYSF 678
 DB 605 NGNEKNKASISIFQKIVAEFEELKVLPEVDSAAALDGSAGVFNRTICRSYPIYKF 664
 QY 679 VREELGVKARGDVPFGKQVTTGSVNSRIYEAISGR 717
 DB 665 VREELG-----AEYLTGEKVRSPGECDDKVFALICEGKI 698
 RESULT 12
 PALL PETCR STANDARD; PRT; 716 AA.
 AC P2481;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
 GN PALL.
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
 OC Apium clade; Petroselinum.
 OX NCBI_TaxID=4043;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89356635; PubMed=2767049;
 RA Lois R., Dietrich A., Hahlbrock K., Schulz M.;
 RT "A phenylalanine ammonia-lyase gene from parsley: structure,
 RT regulation and identification of elicitor and light responsive
 RT cis-acting elements";
 RL EMBO J. 8:1641-1648 (1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94326930; PubMed=8050576;
 RA Schuster B., Peley J.;
 RT "Serine-202 is the putative precursor of the active site
 RT dehydroalanine of phenylalanine ammonia lyase. Site-directed
 RT mutagenesis studies on the enzyme from parsley (Petroselinum crispum
 RT L.)";
 RL FEBS Lett. 349:252-254 (1994).

CC -I- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC -I- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC -I- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -I- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (by similarity).
 CC -I- SIMILARITY: Belongs to the PAL / histidase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL: X15473; CAA33500.1; -
 CC EMBL: X16772; CAA34725.1; -
 CC EMBL: Y07654; CAA8938.1; -
 CC PIR: S04463; S04463.
 CC HSSP: P21310; 188P.
 CC InterPro: IPR008948; L-Asparagine-lyase.
 CC InterPro: IPR001106; Phe/His NH3lyase.
 CC InterPro: IPR005922; Phe_am_lyase.
 CC Pfam: PF00221; PAL; 1.
 CC TIGRfam: TIGR01226; phe_am_lyase; 1.
 CC PROSITE: PS00489; PAL_HISTIDASE; 1.
 CC KMW: Phenylpropanoid metabolism; Multigene family.
 CC FT CROSSLINK 202 204 5-imidazolone (Ala-Gly)
 CC FT (by similarity).
 CC FT MOD_RES 203 203 DHA (2,3-DIHYDROALANINE)
 CC FT (by similarity).
 CC FT MUTAGEN 203 203 S->A: COMPLETE LOSS OF ACTIVITY.
 CC FT 210 210 S->A: NO LOSS OF ACTIVITY.
 CC SQ SEQUENCE 716 AA; 77828 MW; 23D82FALACFBBA4 CRC64;
 Query Match 28.1%; Score 946.5; DB 1; Length 716;
 Best Local Similarity 36.7%; Pred. No. 1e-65; Indels 43; Gaps 18;
 Matches 254; Conservative 110; Mismatches 286;
 QY 49 TGLDIVEXXADPXTDXELDGYSLTGLGVYG-PARKGXVRYXDSERXKIDKSVFE 107
 DB 40 SHLDEVKQWVAB-YRKPVYKLGELTISQVAAISADSGCVYELSEARAGVXKSPW 98
 QY 108 LRXQLXNSV--YGVTTGGSGADRTEDATSLQKALLHQLGCVLPFSKXSFYLGLEN 165
 DB 99 VMDSVNKKGTDSYGVTTGGATSHRTKQGGALQKELIRFNAGI-----FNGSDN 149
 QY 166 SLPLEVVGAMTIRVNSLTRGSAVRLVLEALTNFNLHGITPVPJRGTISAGDLSPL 225
 DB 156 TLPHSATTAAMVAVNLITLQGYSGIRFEILKATKFNQNTTCLPARGTITSGDVLPL 209
 QY 226 STIAAATIGHPDSKHYVHHEGKIXKARALFGLER--VVLGREGJGLVNGTAVSA 283
 DB 210 SYIAQLTGRBNSXA-VGPTGV--ILSPFEAFKLAGVEGFELOPEGLALVNGTAVS 266
 QY 284 SWATLALDHANLISLSQALPTLVEAMVAGAGSFHPLDVTREPHPTOLEVARNIRTL 343
 DB 267 GKASWVLFENILAVLAEVNSAIFAEMQKR-EFTDHLTHKHKHNGQLEAAIWEHL 325
 QY 344 EGSKFAVHEEEVAVKDEGLRDPRPLTSPQMLGPLVSDMTHAVALSLDAGSTTD 403
 DB 326 DGSAY-VAAQKLIEMDPLQKPKDRYALRTSPQMLGPOLE-EVIRISTMIREINSVND 383
 QY 404 NPLIDVENKCTHGGNFQASAVNTMEKTRMLALIGKLNFTQTEMNAGNRGIPSC 463
 DB 384 NPLIDVENKCTHGGNFQASAVNTMEKTRMLALIGKLNFTQTEMNAGNRGIPSC 443
 QY 464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAENQAVNSIALISARRTAE 522


```

Db          444  DGGRRPISDYGEFKAEIETAMASVCEGELFLANPVNTNVOQAEOHQNDVNSLGLISRTKSE 503
2y          523  ANDVLSLLAHILICVQLQAVLRAMEPEPFKQFPXXXXXLLHQFGKXKTXXXXXXELX 582
Db          504  AVEILIKMSTFVLVGLCGAIDLRHLEENLKSTVKNVTSVAKRPLTWNGVGLHPSRFCE 553
2y          583  KXXXXKXRLKEGTNSYDLEP--RWHDAFSXATGTVEKLSXXXXXVSLAAVANAKVAXA 640
Db          564  KDLARVDR-ETFPAYIDDPCSNATYPLMQKRLQRLVBEHALNGNDENMLSPISFGKATP 622
2y          641  EKAIS--LTRVYKXXFWXAPSSSSPALXYSLP-----RTFVLYSFVRELGVKARGDV 692
Db          623  EDELKALLPKVEYSA-RAALSNGNPAI-----PNRIECSRYPFLYKFKELGT-----EY 672
2y          693  FLGKQVITIGSVSRVYEAISGRTNXLVVMVL 725
Db          673  LTGEKVTSPGEEFEKVFYAMSKGEIIDLLECL 705

RESULT 13
PAL2_PETCR
ID      PAL2_PETCR      STANDARD:      PRT:      716 AA.
AC      P45728;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Phenylalanine ammonia-lyase 2 (EC 4.3.1.15).
GN      PAL2.
OS      Petroselinum crispum (Parsley) (Petroselinum hortense).
OC      Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC      Spermatophyta, Magnoliophyta; eudicotyledons, core eudicots, asterids;
OC      campanulids, Apiales; Apiaceae; Apioideae; apioid superclade;
OC      Apium clade; Petroselinum.
OX      NCBI_TaxID=4043;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9501041, PubMed=7925471;
RA      Appert C., Logemann E., Hanbrock K., Schmid J., Amrhein N.;
RT      "Structural and catalytic properties of the four phenylalanine
RT      ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL      Eur. J. Biochem. 225:491-499 (1994).
CC      -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC      first reaction in the biosynthesis from L-phenylalanine of a wide
CC      variety of natural products based on the phenylpropane skeleton.
CC      -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC      -1- PATHWAY: Phenylpropanoid biosynthesis, first step.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC      which is formed autocatalytically by cyclization and dehydration
CC      of residues Ala-Ser-Gly (By similarity).
CC      -1- SIMILARITY: Belongs to the PAL / histidase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      CC use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X81158; CA57056.1; -.
DR      PIR: S48725; S48725.
DR      HSSP: P21310; 1B8F.
DR      InterPro: IPR008948; L-Asparagine-like.
DR      InterPro: IPR001106; Phe/His NH3Lase.
DR      InterPro: IPR005922; Phe_am-Lyase.
DR      Pfam: PF00221; PAL; 1.
DR      TIGRFAMs: TIGR01226; phe_am_lyase; 1.
DR      PROSITE: PS00488; PAL_HISTIDASE; 1.
KM      Lyase: Phenylpropanoid metabolism; Multigene family.
FT      CROSSLINK 202 204 5-imidazolinone (Ala-Gly)
FT      (By similarity)
FT      MOD_RES 203 203 DHA (2,3-DIHYDROALANINE)

```

ST	SEQUENCE	716 AA, 77964 MW; F74D7BD861BEE66 CRC64;	(BY SIMILARITY)
Query Match	28.1%; Score 946.5; DB 1; Length 716;		
Best Local Similarity	35.5%; Pred. No. 1e-65;		
Matches	254; Conservative 110; Mismatches 264; Indels 87; Gaps 19;		
QY	49 TOLDIYEXKALPTXDDXDELDGYSLTLDGVWG-AARKGRYVXNDSEIRKIDKSEVF	107	
DB	40 SHLDEYKKNVAE-YKRPVYKLGCELTITISQVAAISANDSGGVTVELSEARAGVAKASDW	98	
QY	108 LRXYLXNSV--YGVTTGFGSGADTRTEDAISLQKALLHQLCGVLPFSKDSYLRGLGN	165	
DB	99 VMSNMKKGDSGVGTTGFGATSHRRTKGGALQCELRFLNMAI-----FGAGSDN	149	
QY	166 SLPLEYVRGAMTIRVNSLTPRGSAVRVLVEALTNFINHGITPILVLRGITSASGDLSP	225	
DB	150 TLPHSAITRAAMLYRINTLLQSGSIRPEILEAIKFLNQGMITCCLLRGTTIASGDLVPL	209	
QY	226 SYTAAAITGHPDSKYVXVHEGEXKIMKAREALFGLER--VTLGRGELGVNGTAVSA	283	
DB	210 SYIAGLITRPFNSKA-VGPTGV--ILSPEEAFKLAGVEGGEFELQKEGIALVNGTAVS	266	
QY	284 SMATTLIHQAHLISLSQALTLYEAVNVAAGSFFPLDVTVRPHPTQIEVARNITLL	343	
DB	267 GMSMWLFEANILAVLAEMSAIRFEVNGSR-FETDHLNKLKHHPQGEAAMIMHTL	325	
QY	344 EGSAFVAHHEEYKVKDDEGILLRQDYPRTSPQWLCPLVSDMIIAHAVLSLEAGSTTD	403	
DB	326 DGSAY-VKAQKLEHMDPLQKPKQDRAVLRTPSPQMLPQI-EVIRSTKYIEREINSVD	383	
QY	404 NPLIDVANKCTHHGGFQMSAVXNMEXTRLALLIGKLPDTULTKMGMRGLPSC	463	
DB	384 NPLIDVSRNKAIHGGFQGTPIGMSMDTRIALAIGKLPAPQSELVANDFYNGGLPSNL	443	
QY	464 A-AEDPSLSYCKGLDIAAATYSELGHANPVYTHVQAPAMGNOAVNSIALISAEITAE	522	
DB	444 SGGNPELDGFEFGAETIAVASCSELQFLANPVTNHVQSAHQNDVNSLGLISRKTS	503	
QY	523 ANDVLSLLATHTLYCYLVQVLDRAAEFEFKQKPYKXXLLXCHF-----GXKXTXXXX	577	
DB	504 AVELIKLMSTFLVGLCOALDIRLHEMLKSTVKNYVSSVAKRVLTVNGVNGELHPSRFE	563	
QY	578 XELKXKXXKX-----LXKRLBQ-----NSYDLEPRMDDAFSXAATGVV	616	
DB	564 KDILRFIDRRIYFAITDDPCSATYPLMQKLTQVLEHALKNGDERNMNTSIFQKIATFE	623	
QY	617 EXLSSXXXXXVSLAANAMKVAKAKAISLRXYKXXKWKAPSSSALXLYBP-----	670	
DB	624 DELK-----ALPKEVESAR-----PALESGNPAT---PNRIEBC	655	
QY	671 RTRYLVSFVEEELGVKARGDVFGLQGEWTIGSNVRIYEAIKSGRINXVLYVKL	725	
DB	656 KSYPLLYFVAKELGI-----EYLTGKNTSPSEEFDKYFIMSGEIIIDPLECL	705	
RESULT 14			
PAL1	SOYBN	STANDARD;	PRT; 713 AA.
AC	P27951:		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Phenylalanine ammonia-lyase I (EC 4.3.1.5).		
GN	PAL1.		
OS	Glycine max (Soybean).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid 1; Fabales; Papilionaceae; Papilionoideae; Phaseoleae; Glycine.		
OX	NCBI_TaxID=3847;		
RN	[1]		
RP	SEQUENCE FROM N. A.		
RC	STRAIN=cv. T225; TISSUE=leaf;		

```

EX MEDLINE=92190550; PubMed=1799682;
RA Frank R.L., Vodka L.O.;
RT "Sequence and structure of a phenylalanine ammonia-lyase gene from
RT Glycine max."
RL DNA Seq. 1:335-346(1991).
CC
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylcyclohexanone-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC -----
DR EMBL; X52953; CA37129.1; -.
DR PIR; S22991; S22991.
DR HSDB; P21310; 1B8F.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His_NH1Lyase.
DR InterPro; IPR005922; Phe_am_Lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
DR Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLINK 199 201 5-imidazolinone (Ala-Gly)
FT MOD_RES 200 200 DHA (2,3-DIHYDROALANINE)
FT FT (By similarity).
SQ SEQUENCE 713 AA; 77744 MW; 9D71EF1CC230216A CRC64;

Query Match 28.1%; Score 946; DB 1; Length 713;
Best Local Similarity 36.6%; Pred. No. 1,1e-65;
Matches 271; Conservative 105; Mismatches 293; Indels 72; Gaps 23;

QY 14 NGXXNGXHAAXXAXXXXXXXXAXAGSXLPTTXXTQDIVEXXLADPYTDXDXELDYS 73
DB 5 NGHONGSFCGSTAKGNNDPLNWGAAA---EAMKSHLDEYKRWVAE-YRKPVRLGGERT 59
QY 74 LITGDVVGAAKRGKRYKVDSEITXKIDKSYEPIRXOLXNV--YGTTFGFGSADTRT 111
DB 60 LITQVAAVAGHDGVAELSESAREGVKASSEWMSNMNGTDSYGTTFGFGATSHRR 119
QY 132 EPAISLQKALIEHOLCGVLPSTXDSFPLGRLGNS--LPLEVGAMTIRVNSITRGSA 189
DB 120 KCGGALQKELIRFINAGI-----FNGTGESHTLPHRATGAALVAVINTLLQSYG 170
QY 190 VALVLEALTNFLNHGITPIYVLRGTIASGDLSPSTIAAATGHPDSKRVYHHEGXEK 249
DB 171 IFFELIEALTKLNNNVTPCDDLRTGTIASGDLVPLSYAIDLTRGNSXA-VQPSG-E 227
QY 250 IXXAREBALFGL--EPVVLGPKEGGLVNGTAAVSASAMTLADHMLSLTSQLTALT 307
DB 228 VINAKEAFELASINSEFELQPKBELAVNGTAVSGSLASVPLFEANILAVLSVSAIF 287
QY 308 VEAATGAGSGSPHPHDVTRPHAPQIEVARNIRFLLEGSXFAYV---HEEYKVDDEG 363
DB 288 AEVWQGP-EFTDHTLTKLKHHPQIEAAVMEIILDGSSYMKAKKHLIDLPQKP--- 343
QY 364 ILROORYPLRTSPQGLPVSDMTAAHAAVLSLENG-QSTTDNPLIDVENKXTHGNGFOA 422
DB 344 --KORRYALRTSPQWLPPIE--VIRSTSTSEINSVDNPLIDVSRKALHGNFQ 399
QY 423 SAVXNTEKRLALALIGKLNFTQLTEMLNAGMNRGDPGL-AAEDPSLSYHCKGLDIAA 481

```

CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3) .
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-midazole-5-one (Mio),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

	CC	
DR	EMBL; ACC09400; AAF02809.1;	-.
DR	HSSP: P21310; IBAF.	
DR	InterPro: IPRO08968; L-Aspartase-like.	
DR	InterPro: IPR001106; Phe/His_NH3lyase.	
DR	InterPro: IPR005922; Phe_am_lyase.	
DR	Pfam: PF00221; PAL; 1.	
DR	TIGRfams: TIGR01226; phe-am lyase; 1.	
DR	PROSITE: PS00488; PAI_HISTIDASE.1.	
Kw	Lysine; Phenylpropanoid metabolism; Multigene family;	
Ft	CrossLink	194 196 (By similarity).
Ft		DHA (2,3-DIHYDROALANINE)
Ft		(BY SIMILARITY).
FT	MOD_RES	195 195
SQ	SEQUENCE	707 AA; 76919 MW; B67743C9A1B50B9B CRC4;

Query Match	28.0%;	Score 943.5;	DB 1;	Length 707;
Best Local Similarity	37.0%;	Pred. No. 1.7e-65;		
Matches 252;	Conservative 98;	Mismatches 282;	Indels 49;	Gaps 16;

```

QY 65 DXXEILDGVSILTLGDVYGAAKRGKRYKRVYDSDDEIIXKIDKSVELRLXQOLNSV--YGVTTG 122
Db 45 EAVKLGELETLTIGYAAAVARGGSGSYTELAEAAAGYKASSEVMMSNMNRGDSYGVTTG 104
QY 123 FGGSAADRTEDIASIQKALIEHOLGVLPTYSXDSFXLGRG-LENSIPLEVYRGAMTIRYN 181
Db 105 FGAISRFRKQGGALQONELIRFLNAGI-----FGGAADTSHTLPKPTRAAMLVRYN 157
QY 182 SLTRGSAATVLEVALNPNLNGHTFVPLRGYTSASGDLPSLSTYAAATIGHPDSKYH 241
Db 158 TLLQGSVSGIRPELLEKTYLKLNLHETLPCLPRGTTTASGDLVPLSTIAGLIGRPSKRA- 216
QY 242 VXEHEKEXIMXAREAIALFGLRPVY-LGPKGELGVNATVAVSASMTLALHDAHMLSLIS 300
Db 217 VGRSG--EFTLTSSEAFKLAGVSSPELQPKGGLVNGTAVGSGLASVLPDANILAVLS 274
QY 301 QALITALTVANMGHAGSHFPLPHDTNRHPQIOEVANINITLLEGSKFA---VHHEEV 356
Db 275 EVMSAFMAFWMGKR-EFTDLTLTKLKHPCQIIPAAIMHILIDGSSVYKEAQLHENDP 333
QY 357 KYVDEBGLIROPYPLRTSPQWGLGVSDMHAHVAVLSBAGOSTDNPPLIDENKXTXH 416
Db 334 LQRP-----KQRYALRTSPQWLGQI-EVIRATKXIEEINSVYNDPLIDVSRKALH 387
QY 417 GGNFQASAVXNTEKTRIALALIGKALNTQJTEMILAGNNGRLPSCIA-AEDPSLSYHCK 475
Db 388 GGNFGQPTGIVANDNRRLIASIGKLMAFQSESLVNDFYNNGJLPSNLISGRNPSLDYGRX 447
QY 476 GJLIDAAATSEIGHLANPVTHHQPAEMGNQVNSLALISARPLEANDVUSLILATL 535
Db 448 GAELIAASYSSELQFLANPVINHVSAGHQNDVNSGLLSSRKTAENDILKLMSTTL 507
QY 536 YCULIADTVIRAMEFEFKQFXEXXXXXLXOHFGXXXTLXXXXXXELXKVKXKLXKLEOT 595
Db 508 VALCQAVDLRHLEENIKRAVSVAGVAKRVLTVGANGELHSPNFTERDVLQVDR-EVY 566
QY 596 NSYDLEB-----RWEDAFSYATGTVEYKLSXXXXXVSLAAVANAKVLAEXAKLS 645

```

```

Db      567 FSVADDPCLTYPLMOKLRLHVLVDALADPERENASATVPHKGAPE-----AEIKL 620
Qy      646 ITRXV-RXXFWAPSSSPALXYLSPPRNVYSFVREELGVKARGGVPLCKQVITGSN 704
Db      621 LPEVEVERVEVEEGTSATANRIKCRSYPYRFVDELNT-----ELTGENVRSGEE 675
Qy      705 VSRIVEAIRSGRINXVLYVMT 725
Db      676 FDKVFLAISGKLIDPLECL 696

```

Search completed: September 9, 2004, 10:04:42
 Job time : 19.5685 secs

```
Search completed: September 9, 2004, 10:04:42
Job time : 19.5685 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:57:28 ; Search time 88.3651 Seconds
(without alignments)
2592.269 Million cell updates/sec

Title: US-09-939-408a-21

Perfect score: 3365
Sequence: 1 NAFPSDSTATSXANGXNGX.....RIYAIKSGRLNVLVMLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	969	28.8	725	10 Q94ANI	Q94ani arabidopsis
2	967	28.7	711	10 Q8W2B4	Q8w2b4 lactuca sat
3	965.5	28.7	707	10 Q7X7G3	Q7x7g3 cryza sativ
4	963	28.6	671	10 Q84VRO	Q84vro cryza sativ
5	958.5	28.5	703	10 Q8VXG7	Q8vxg7 zea mays (m
6	955	28.4	711	10 Q9AXI5	Q9axi5 pharbitis n
7	955	28.4	730	10 Q9W567	Q9w567 rubus idaeu
8	954	28.4	708	10 Q94ENO	Q94enu rehmannia g
9	944.5	28.1	681	10 Q8RUZ3	Q8ruz3 pinus sylve
10	944.5	28.1	715	10 Q85ZS1	Q85zs1 daucus caro
11	943.5	28.0	681	10 Q8SAS6	Q8sas6 pinus sylve
12	942.5	28.0	716	10 Q8WAX1	Q8wax1 catharantu
13	941.5	28.0	681	10 Q8RW49	Q8rw49 pinus sylve
14	940	27.9	717	10 Q8RWP4	Q8rwp4 arabidopsis
15	939	27.9	713	10 Q8GZR8	Q8gzz8 lactuca sat
16	937.5	27.9	681	10 Q8SAS7	Q8sas7 pinus sylve

17	931	27.7	710	10 Q94C45	Q94c45 manihot esc
18	929.5	27.6	716	10 Q94IN7	Q94in7 agastache r
19	926.5	27.5	721	10 Q9XFX5	Q9xfx5 citrus clem
20	922	27.4	687	10 Q94FP0	Q94fp0 manihot esc
21	919.5	27.3	727	10 Q7XJC3	Q7xjc3 pinus pinas
22	917.5	27.3	727	10 Q7XJC4	Q7xjc4 pinus pinas
23	915.5	27.2	714	10 Q7XCM8	Q7xcm8 cryza sativ
24	914	27.2	711	10 Q8H6V5	Q8h6v5 populus tre
25	911.5	27.1	714	10 Q8H6V6	Q8h6v6 populus tre
26	910	27.0	717	10 Q8H6W0	Q8h6w0 coffea cane
27	904.5	26.9	712	10 Q94F89	Q94f89 manihot esc
28	900	26.7	655	10 Q9XG83	Q9xg83 vigna ungu
29	898.5	26.7	549	10 Q04876	Q04876 hordeum vul
30	895.5	26.6	718	10 Q9XFX6	Q9xfx6 citrus clem
31	893	26.5	703	10 Q84I72	Q84i72 phalaenopsi
32	893	26.5	715	10 Q24286	Q24286 populus kit
33	892.5	26.5	619	10 Q39545	Q39545 cucumis mel
34	887	26.4	619	10 Q8H6V9	Q8h6v9 coffea cane
35	868.5	25.8	618	10 Q9FS00	Q9fs00 dianthus ca
36	763	22.7	396	10 Q8WVP6	Q8wvp6 beta vulgar
37	736.5	21.9	544	10 Q9FT38	Q9ft38 trifolium ae
38	735	21.8	532	10 Q81647	Q81647 capicum ch
39	717	21.3	380	10 Q84U27	Q84u27 ginkgo bil
40	715	21.2	362	10 Q9FU17	Q9fu17 rubus idaeu
41	703	20.9	363	10 Q9FUT8	Q9fut8 rubus idaeu
42	612.5	18.2	497	10 Q04869	Q04869 hordeum vul
43	604.5	18.0	475	10 Q04875	Q04875 hordeum vul
44	600.5	17.8	436	10 Q9LEP1	Q9lep1 betula verr
45	596.5	17.7	435	10 Q9M3U1	Q9m3u1 betula verr

ALIGNMENTS

RESULT 1

Q94ANI PRELIMINARY: PRT: 725 AA.

AC Q94ANI

ID Q94ANI

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Phenylalanine ammonia lyase (Hypothetical protein) (EC 4.3.1.5).

GN At2G37040 OR PAL1.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosoid II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN

RP SEQUENCE FROM N.A.

RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,

RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,

RA Yamamura Y., Yu G., Yu S., Bower L., Carninci P., Chen H., Cheuk R.,

RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,

RA Kawai J., Kim C., Kosena E., Lam B., Lin J., Meyers M.C., Miranda M.,

RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,

RA Shim P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,

RA Ecker J.R., Theologis A.

RT "Full length cDNA of gene TlJ8-22/At2G37040 (GI:5598547).";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,

RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,

RA Dang J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,

RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,

RA Satou M., Seki M., Shim P., Tang C.C., Toriumi M., Wong C., Wu H.C.,

RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,

RA Davis R.W.,

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etku P., Lee J.M.,

RA Tortum M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shim P.,
 RA Southwick A., Davis R.M., Becker U.R., Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cochran F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.,
 RT "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
 RT Characterization of Recombinant Proteins and Knock-out Mutant
 RT Characterization."
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY045919; AK76593.1; -
 DR EMBL; BT003330; AA029949.1; -
 DR EMBL; AY079363; AL85094.1; -
 DR EMBL; AY303128; AAP59438.1; -
 DR PIR; G84787; G84787.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016841; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His_NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; phe am lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KM Hypothetical protein; lyase.
 SQ SEQUENCE 725 AA; 78725 MW; 026263B2DEFE9CE CRC64;

Query Match 28.8%; Score 969; DB 10; Length 725;
 Best Local Similarity 35.5%; Pred. No. 8.8e-68;
 Matches 258; Conservative 120; Mismatches 257; Indels 92; Gaps 21;

QY 40 GSKPTTXXTQLDIVEXKXLDADPTDXKXLDGYSITLGDVYGAARKGXRYXSDERX 99
 DB 39 GAABEQMKSHLEVRKVAE-FRKPVNVLGGETLTIGQVAAISITGNSVKELSETARA 97
 QY 100 KIDSVYELRXQLXNSV-YGVTTFGSGADRTEDAIISQKALDEHQLCGVLPTSDXSF 157
 DB 98 GVAASSDMVMSNMKGTDSYGVTTFGATSHRRTKNGVALQKELRFLMAIGFSTKRT- 156
 QY 158 XLGRGLNSLPLEVRGAMTIRVNSLTRGSAVRLVLEFALTNFNGITPVLPGTIS 217
 DB 157 -----SHTLPHSATRAAMLVIRINTLQGFSGIRFEILLEATISFLNNITPSPLRGIT 210
 QY 218 ASGDLSPSTIAAATGHPDSKVHVXHEGEXKIMAREALFGLER-VVLGPRKGL 275
 DB 211 ASGDLVPLSTIAGLIGRPSKA-TOPNG-EALTAERFKLAGISSGFPELPKREGIAL 267
 QY 276 VNGTAVSASWATLADHMLLSLQALTAITVEAMVGAGSFHPLHDVTRPHPTQIEV 335
 DB 268 VNGTAVSGMAWVLEFETVVLSTAEILSAVFAVMSGKP-EFTDHLTHRLKHPGQIEA 326
 QY 336 ARIRITLLGSXF-----AVHHEEVYKXKDEGILADQRYPLRTSQWIGPLVSDMIANA 391
 DB 327 AALMEHILDGSSWMLKLAQKLEHNDLPQKP-----KQDRYALRTSQWIGPQIE-VIRYA 379
 QY 392 VLSLEAG-OSTEDNPLDIVENKXTHGNGFQASAVXNTMEKRLTALIGLNFQLTLEM 450
 DB 380 TKSIEREINSVNDNPLDIVSRKKAHGNFGQTPGVGMDNTRLAIAIGLMAFQSEL 439
 QY 451 LNAQMGRLPSCI-AAEDPSLSYHCKGLDIAAAATSETGLANPVTTHVPAEMQAY 509
 DB 440 VNFYNNGLPSNLTARNSLDYGFGEAIAWASCSLQYLANPVSHVQSAQHQDV 499
 QY 510 NSIALISARTAEANDVLSLLATHLYCVLQAVDRAMEFEFKQFXXXXLXQFGX 569
 DB 500 NSIGLISRSKRTSAVDILKLMSTTFLVAILCQAVDRLHLENLQ-----543
 QY 570 XTXXXXXXELXXKX'-----XKXKXRL-BQNSYDLEPRMHDAFSXAT 612

DB 544 --TVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLKIVDRQVITYADDP-----CSAT 595
 QY 613 GTVEXLSSXXXXXVSLAAVNA-----MKVAAEKAIS-LTRXVXXFWXAPS 659
 DB 596 YPLIQKLRQ---VIVDHALINGSESKNAVTSIFPKHGAFFEEELKAVLPKEVEARAAYDN 652
 QY 660 SSSPALXVSP-RTVVLVSFPREELGVAKRGDVLGQVETIGTSNNSRIYEAIKGRIN 718
 DB 653 GTSALPNRIKCRSRYPLEFRVRELDGT-----ELLTGKVTSPDEEPDKVFTALCEKII 707
 QY 719 XVLVKML 725
 DB 708 DPMVECL 714

RESULT 2
 Q8W2E4
 ID Q8W2E4 PRELIMINARY; PRT; 711 AA.
 AC Q8W2E4;
 DT 01-MAR-2002 (TREMREL. 20, Created)
 DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase.
 OS Lactuca sativa (Garden lettuce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Asterales; Asteraceae; Cichorioideae; Cichoriaceae;
 OC Lactuca.
 OC NCBI_TaxID=4236;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Campos-Vargas R., Nonogaki H., Suslow T., Salveit M.,
 RT "Characterization of phenylalanine ammonia-lyase (PAL) gene in wounded
 RT lettuce leaf tissue."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF299330; AAL55242.1; -
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016841; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His_NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; phe am lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KM lyase.
 SQ SEQUENCE 711 AA; 77363 MW; 3120AB2877C29513 CRC64;

Query Match 28.7%; Score 967; DB 10; Length 711;
 Best Local Similarity 36.9%; Pred. No. 1.2e-67;
 Matches 263; Conservative 108; Mismatches 262; Indels 80; Gaps 20;

QY 49 TQLDIVEXKXLDADPTDXKXLDGYSITLGDVYV-AARKGXRYVXSDERIRKIDKSYE 106
 DB 32 SHLDVKKXVAE-FRKPVYKLGGETLTVSQVAGIAAANDSDTYKVELSEARAGVASSD 90
 QY 107 FLRXQLXNSV-YGVTTFGSGADRTEDAIISQKALLEHQLCGVLPTSDXSFYLRGDE 164
 DB 91 WMSNMKGTDSYGVTTFGATSHRRTKNGVALQKELIRFLNNGI-----FNGTRE 141
 QY 165 NS-LPLEVRGAMTIRVNSLTRGSAVRLVLEFALTNFNGITPVLPGTISAGDI 222
 DB 142 TSHLPHSATRAAMLVIRINTLQGFSGIRFEILLEATIKFPLNNITPCLPLRGITIASGDL 201
 QY 223 SPLSTIAAATGHPDSKVHVXHEGEXKIMAREALFGLER-VVLGPRKGLVNGTA 280
 DB 202 VPLSTIAGLIGRPSKA-VQPTG-EVLAERKAFPAAGVGGFPELPKREGIALVNGTA 258
 QY 281 VSASWATLADHMLLSLQALTAITVEAMVGAGSFHPLHDVTRPHPTQIEVANR 340

```

Db      259 VSSGASWFLPAPNALALISEVLSAIFAEVMOGRK-ETDMLTKLKHHPQITAAIME 317
QY      341 TLLEGSFAVHNEEEVKYKDEGILRQDRYPLRTSPQWLGVLSDMTAAHVAVSLERAGQS 400
Db      318 YLIDGSDY-VKAAQKHEMDPLQKPKDRVALRTSPQWLGFOI-EVIRSTRKMIEREINS 375
QY      401 ITDNPILDVENKXTHHGNFQASAVXNMTEKRLALALIGKLNFTOLTLEMNAGMNGLP 460
Db      376 VNDNPILDVSRNKAHLHGNGFQSTPIGVSMNTRLAALIGKLMFAQSELVNDYRNGLP 435
QY      461 SCIA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISAR 519
Db      436 SNLSGGRNPSLDYGFGRGELTAAASCESELQFLANPVTNHVQSAEGHNDVNSLGLISAR 495
QY      520 TAEANDVLSLLATHLYCVLQAVDLRAMEFEFKQFXXXXLXQH- -GXHTXX 574
Db      496 TAEAVDILKMSSTYVALQSIDLRHLENNKSTVKNVTQVQAKVLTWGVNEELHPSR 555
QY      575 XXXXELXXVXXX- -LXKRLQCT- -NSYDLPRWHDAPFSXATG 613
Db      556 FCKKDLRLRVDEYPAATIDVCSGYPLMCKLRGVLDHALNNGETEKNTNITLFOKIA 615
QY      614 TVEXLSXXXXXVSLAANAMKVAAXAEKALSLTXVXXXTWAPSSSPALXLSPTTR 673
Db      616 TFEELK- -VLLPKEVGRIAYENDTLSPRIK- -ACRSY 653
QY      674 VLYSFVREELGVKARGDVEFKQEVTL- -GSMVSYIEAKISGRINXVLYM 725
Db      654 PLRFRVRELG- -RG- -FLTGEKYSFGEFDFVFMCKGQILDLECL 700

RESULT 3
QYX7G3 PRELIMINARY; PRT; 707 AA.
ID QYX7G3
AC QYX7G3;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE OSJNB0073E02.12 protein (OSJNB0061C13.1 protein).
GN OSJNB0073E02.12 OR OSJNB0061C13.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Lu Y.L., Mu Z., Chen L., Fan D.L., Meng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.T., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Wang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.F., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.U., Chen S.T., Ni L., Zhu F.H., Hong G.F.,
RA Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL731616; CA05452.1; -
DR EMBL; AL731629; CA05619.1; -
DR SEQUENCE 707 AA; 76038 MW; 841E4C6D077B1AA CRC64;

Query Match 28.7%; Score 965.5; DB 10; Length 707;
Best Local Similarity 37.1%; Pred. No. 1.6e-67;
Matches 257; Conservative 122; Mismatches 273; Indels 41; Gaps 20;

49 TOLDIVEXXADXTDXKELDGYSLTLGPV-VGAARGRVYRVXSDPEIKXIDKVE 106
Db 29 SHLDVVRKMAQ-SRENVVKEISSSLRVGVAVASAAKASGVVELDEARPRKASE 87
QY 107 FLRXQLXN-SYGVTTGGGASDPTEDALISQKALLHQJLGVLPSTXSDFXLGRGLE 164
Db 88 WILNCIAHGDDIGVTTGGTSHRRTKQGOALQVELRLHNLNAGIFGNGSDG----- 139

```

```

QY      166 NSLPLEVYRGAMTIRVNSLFRGSAVRLVLEALTNTFNHNGITPIVPLRGSTISAGDLP 224
Db      140 NSLSEVSRRAALVAINTLLOGISGIREILFALTKLINTGVSCLRLRGITTSGLVP 199
QY      225 LSYIAAITGHDSKVYHHEGKEXKXKARALALFGLP- -VLGREGIGLVNGTAVS 282
Db      200 LSYIGLITGRNQA-VTVG- -KVDAEAFAKIAGGFFRLEKEGIALVNGTSVG 256
QY      283 ASMATLALHDAMLSLSQALTALTVEAMVGHASFFPLHDVTRPPTQEVARNRTL 342
Db      257 SALAMVLYDANVALVSEVLSAVFCEVNGKPEYTHLTKL-KHHGISTEAAIAIHEH 315
QY      343 LSGSFAVHNEEEVKYKDEGIL- -RQDRYPLRTSPQWLGVLSDMTAAHVAVSLERAG-Q 399
Db      316 LSGSFAVHNAQ- -KNEVVDPLKPKQDRVALRTSPQWLGFOI-EVIRA-ATKSIEREVN 370
QY      400 STTDNPILDVENKXTHHGNFQASAVXNMTEKRLALALIGKLNFTOLTLEMNAGMNGRL 459
Db      371 SYNDNPVIDVHRGKALHGNGFQSTPIGVSMNTRLAALIGKLMFAQSELVNEFYNNGL 430
QY      460 PSCIA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISAR 518
Db      431 TSNLGSRRPISLDYGFGRGELTAAASCESELQFLANPVTNHVQSAEGHNDVNSLGLISAR 490
QY      519 TAEANDVLSLLATHLYCVLQAVDLRAMEFEFKQFXXXXLXQHFGKXXTXXXXXX 578
Db      491 TAEAVDILKMSSTYVALQAVDLRHLLENNKSAVNCVTTAKRVLTTPAGGLHSA 550
QY      579 ELXXVXXKXKRLQNTSYDLEP- -RWHDAFSXATGVEXLSXXXXXVSLAANAMK 636
Db      551 RPSERKALTAIDR-EAVSYADDPSCANYPLMTKIRAVLVHALANGAECDGSSVPSK 609
QY      637 VAXAEKAI-SLTRYVXXFWXAPSSSPALXYL-SPTRYLYSFVREELGVKARGDVF 693
Db      610 ITAFEEELREALPREMEARVAFETGTAIPTRIKESRSPLYRFRVRELGC-----VY 663
QY      694 L-GKQEVTTGSMVSRITYEAKISGRINXVLYM 725
Db      664 LTGEKLSFGEBCNKVFALISERKLIDPMLCL 696

RESULT 4
Q84VE0 PRELIMINARY; PRT; 671 AA.
ID Q84VE0
AC Q84VE0;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE Phenylalanine ammonia-lyase.
DE Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipomdare;
RA Cooper B., Hutcheson D., Park S., Gumil S., Lugimbunl P., Eliero C.,
RA Goff S., Glazebrook J.,
RA "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT Mediated Regulation of the Cell Cycle."
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY224546; AA072666.1; -
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; Fatty acid ligase activity; IEA.
DR GO; GO:0016841; Fatty acid ligase activity; IEA.
DR GO; GO:0016829; Fatty acid ligase activity; IEA.
DR GO; GO:0009058; P-biosynthesis; IEA.
DR GO; GO:0006559; P-biosynthesis; IEA.
DR InterPro; IPR008948; L-asparagine catabolism; IEA.
DR InterPro; IPR001106; Phe/His NH3 lyase.
DR InterPro; IPR005922; Phe_am_lyase.

```

DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 671 AA; 72277 MW; 23F93B995CE21E60 CRC64;

Query Match 28.6%; Score 963; DB 10; Length 671;
 Best Local Similarity 37.5%; Pred. No. 2.4e-67;
 Matches 253; Conservative 118; Mismatches 263; Indels 40; Gaps 19;

QY 68 ELDSYSLTLDV--VGAARKGRVRYVSDDELRIKIDKVEFLRXOLN--SVYGVTTGF 123
 DB 11 KIEGSSLRVGVAANVAASAKDASGVVLEDEARPRKASSEMILNCIANGDIYVTTIGF 70
 QY 124 GGSADRTEDALISQKLLLEHQLCVLPFSXDSFKLGRLENSLPLEVVRGAMTRVNSL 183
 DB 71 GGISHRTKDGQALQVELLRHLAGLFGNGSDG-----NSLSEVSRAMLVRIINTL 122
 QY 184 TRGSAAVLVLEALTNFLNHGTTPIVPLRGTTISAGDLSPISTYAAATGHPDSKVAVX 243
 DB 123 LQGVSGIRFELIEAITLINTGVSPLRGTTTASGDVPLSTYAGLITGRPNAQ-VT 181
 QY 244 HEKEXIKMARBEAIALFGLEP--VLGPYBGLVNGVAVASAKMTALIDHAMLSTLSQ 301
 DB 182 VDG--KKVDAAEAFKTAGIQQGFFFLPEPKGLAIVNGISVGSALAMVLYDANVLAJSE 239
 QY 302 ALTALTVAMVGAAGSFHPLHDVTRPHPTQIEVANRTITLLEGSXPAVHHEEVKVKMD 361
 DB 240 VLSAVFCEVWNGKPEYTDHILTKL-KHFGSIEAALIMHILAGSAFPHQO--KQNEV 295
 QY 362 EGIIL--RQDEYPLRTSPQWLGPLVSDIMAHAVLSLEAG-QSTTNDPLIDVENKXTHHG 418
 DB 296 DPLKPKQDXYALRTSPQWLGPOI-EVIRA-ATKSIEREVSVDNPNVIDVHRGKALHG 353
 QY 419 NFGASAVNMTKRTALALIGLNTFTQITLMLNAGMRGLPSCLA-AEDPSLTHCKGL 477
 DB 354 NFGSTPGVSMVTRTALIANIGLMPAFQSELVNFYNNGLSNLAGRNSPLDYGFPGT 413
 QY 478 DIAAAVTSLEGLANPVTHVQPAEMGNQAVNSLALISARETAEANDVLSLLATHLYC 537
 DB 414 EIMASVYSELQFLANPVTHNHQSAEQNHQDVNSLIGVSAKRTAAVAILLMGSTIYVA 473
 QY 538 VLDVNDIRAMEFFPKQFPKXXLLXQHFQXXXTXXKXXKXXKXXKXXKXXKXXKXXK 597
 DB 474 LCAVDLRHLEENLKSAVNCVTVAKKVLITGPAGGLHSARFSEKALLTAIDR-EAVYS 532
 QY 598 YDLEP--RWHDAFSXATGTVEXLSXXXXXVSLAANAMKVAAXEKA1--SLTRVYRX 653
 DB 533 YADDPGSAVYPLMTKIRALVVEHALANGAEKDDSSVFSKITAPEEELREALFREVEAA 592
 QY 654 FKXAPSSSPDALXYL-SPTTRVLYSFVREELGVKARGDVPL-GKQEVTTIGSNVSIYEA 711
 DB 593 RVAFFETAPITNRKIESRSFPLVRFVEBELQC-----VYLTKLKSFGEECNKVFLLA 646
 QY 712 IKSGRLNXYLVKML 725
 DB 647 ISERKLIIDPMLBCL 660

RESULT 5

Q8VXG7 PRELIMINARY: PRT: 703 AA.
 AC Q8VXG7;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCD clade; Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4377;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv, Corso;
 RX MEDLINE=9716112;
 RA Rosler J., Krekel F., Amrhein N., Schmid J.;
 RT "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase
 RT activity";
 RL Plant Physiol. 113:175-179(1997).
 DR EMBL: L77912; AAL40137.1;
 DR GO: GO:0005737; Cytoplasm; IEA.
 DR GO: GO:0016211; F:ammonia ligase activity; IEA.
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO: GO:0009058; P:biosynthesis; IEA.
 DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro: IPR008948; L-Asparatase-like.
 DR InterPro: IPR001106; Phe/His NH3lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 703 AA; 74926 MW; 11374FD68516971E CRC64;

Query Match 28.5%; Score 958.5; DB 10; Length 703;
 Best Local Similarity 34.6%; Pred. No. 5.7e-67;
 Matches 255; Conservative 127; Mismatches 242; Indels 113; Gaps 21;

QY 40 GSXLPTTXXTQDIYEXXLDLPKTDXXELDGYSLTLGDV--VGAARKGRVRYVSDDEL 97
 DB 16 GAAMAEAGSHLDEVRVWAQ-ARQPVYKIEGSTRVGVAANVAASAKDASGVVLEDEEA 74
 QY 98 RXXIDKSYEFLRXOLN--SVYGVTTGGSGADRTEDALISQKALLEHQLCVLPFSXD 155
 DB 75 RPRVAKSEWIDLCIAGGDIYGVTTGFGTSNHRITQDGPALQVELLRHLAGIPIGTSD 134
 QY 156 SFYLGRLNSLPLEVVRGAMTRVNSLITRSHAVRLVLEALTNFLNHGTTPIVPLRG 215
 DB 135 G-----HTLSEVTRRAAMLVRIINTLQYSGIRFELIEAITLINTGVSPLRG 186
 QY 216 ISAGDLSPLSTIAALIGHPSKVNHYHEKEXIKMARBEAIALFGLEP--VLGREG 273
 DB 187 ITASGDVPLSTIAGLITGRPNAQ-VTVDRK--VMAAFAFKIAGIEGFFFLNPEGL 243
 QY 274 GLVNSTAVSASAKMTALIDHAMLSTLSQALTALTVAMVGAAGSFHPLHDVTRPHPTQ 333
 DB 244 AIVNGTSVGSALAAVWYDANVLAJSEVLSAVFCEVWNGKPEYTDHILTKL-KHFGS 302
 QY 334 EVARNIRTLLEGSXPAVHHEEVKVKDDGIL--RQDRYPLRTSPQWLGPLVSDIMHAHA 391
 DB 303 EAAALMEHILDGSSM--KQAKTVNELDPLKPKQDXYALRTSPQWLGPOI-EVIRA-A 357
 QY 392 VLSLEAG-QSTTNDPLIDVENKXTHHGNGFOASAVNMTKRTALALIGLNTFTQITLM 450
 DB 358 TKSIEREVSVDNPNVIDVHRGKALHGNQGPIDGSMONALALANIKLMFAFSESL 417
 QY 451 LNAGMTRGLPSCLA-AEDPSLTHCKGLDPAAYTSLEGLANPVTHVQPAEMGNQAV 509
 DB 418 VNEFYNNGLTSLNLAGSRNPSLDYGFKGTETAMASYSELQYLDNPIITNHQSADEHNQDV 477
 QY 510 NSLALISARPAEANDVLSLLATHLYCVQAVDLRAMEFFPKQFPKXXLLXQHFQX 569
 DB 478 NSLGLVSAKRTAEALIDIKLMSSTIYALQAVDLRHELENK----- 520
 QY 570 XXTYXXXXXXELXXVXXKXKRLQNTSYDLEPRMDAFSXAATGVVEKLSXXXXXVSL 629
 DB 521 -----ASYKNTVTVQAKKVLTMNPSGEL-----SSARFSEKELISA-----IDR 559
 QY 630 AAVNAMKVAAXEKA1SLTRVYRXF-----WKXAPSSSP-- 663
 DB 560 EAVFTYEDASASLPLMOKLRAVLVDHALSSGERGAGLRVLDHQVREGABRGAAPOG 619
 QY 664 -----ALXYLSPTTRVLYSFVREELGVKARGDVPL-GKQEVTTIGSNVSI 708


```

Db      620 GGRPRGVAAGTAPVANRIADSGSPFLYFVRBELG-----VLTGEBRLKSPBECNKV 673
QY      709 YEALKSGRLNXLVVKML 725
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      674 FVGISGKLVDEMECL 690

RESULT 6
09AXIS PRELIMINARY; PRT; 711 AA.
ID 09AXIS;
AC 09AXIS;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS Pharbitis nil (Violet) (Japanese morning glory).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=35883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet.
RA Nakazawa A., Nozue M., Yasuda H., Takeba G., Kubo H.,
RT "Expression of phenylalanine ammonia-lyase in Pharbitis nil.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF325496; AAG49585.1; -.
DR HSSP; P21310; 1B8F.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRPFAM; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 711 AA; 77085 MW; C86EB3A0C0969E90 CRC64;

Query Match 28.4%; Score 955; DB 10; Length 711;
Best Local Similarity 36.3%; Pred. No. 1.1e-66;
Matches 259; Conservative 103; Mismatches 268; Indels 84; Gaps 18;

QY 49 TOLDIYEXXLIADPXTDDXXELDGYSTLGDVVGAAKRGKXRVVXSDERFKKIDKSYEFL 108
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 34 SHLDEKRWVAE-FRRPAVKLGERTITVAQVAIAISRONAVTVLSEESRAGVASSDWM 92
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 109 RQQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPFSXDSFXLGRGLE-- 164
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 93 MDSMNKGTDSYGVTTGFGATSHRRTKOGALQKELIRFLNAGI-----FANGTESC 143
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 165 NSLPLEVVGAMTIRNSLTRGHSAVELVLEALNFINHGTTPIVPLRGITASGDLSF 224
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 144 HTTPHSATPAAMLVRLNTLLQSGISGRFELPAITKLNNHNTLPCLPRGHTTISGDLVP 203
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 225 LSYIAAATGHPDSKRVVHEXGKIMKAREALFGLER--VVLGPRGGLGVNGTAVS 282
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 204 LSTIAGLIGRPRSKA-VGPNG--ETLNAEALRLAGVVGGFELQPRGGLAVNGTAVG 260
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 283 ASNATLALHDAMLSLISQALTAALTYEAMVGAAGSFHFLDVTFRPHQTQEVANIRTL 342
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 261 SGASAVLFEANLAVLSVLISAFVENVNGRP-BFTDHLTKLGHGQTEAALMEHI 319
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 343 LGGSAVHEHEEVKCKDEGLRQDRYPLRTSPQWGLGVLSMDHAHNAVLSLGGQSTT 402
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 320 LDSSSY-VRAAQCGHMDPLQKPKQRYALRKSPQWLGQI-EVIRATKXIEEINSVN 377
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 403 DNPDLIVENKXTHGNGFQASAVXNTMEKTRIALALIGKLNFTQLTENLAAAGNGLPSC 462
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

Db      378 DNPFLIDVARSKALHGNGPQGTPIGVSMDSNRLALASIGKLLFAQFSELVANDYNNGLPSN 437
QY      463 L-AAEDPBLSHCGGLDIAAAAYTSELGHLANPYTTHVQAPRMGNQAVNSLALISART 521
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      438 LTAGRNPSLDYGFQGAETAMASYSCELOFLANPYTNVQSBCHQNDVNSLGLSARKTA 497
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      522 EANDVLSLLATHYCVLQAVDLRAMEFEFKQCFPXKXKLLXQ-----H----- 566
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      498 EAVDVLKLMSSTYVALCQALDLRFLEENLRNANVNAVTVQAKRTLTMGANGELHPARFC 557
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      567 -----FGXXXTXXXXXXELKXKXKXKXKRLQETNSYDLPEPRHDAFSATG 613
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      558 EKDLRLVVDREVPAVADPPSANYPPLMQKLRQALVDHALQNGSESKN-----TG 607
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      614 TVEKLSXXXXXXVSLAAVANAKVAAEKALSTLRVYRXXFWXAPSSSPAL--XYSR 671
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      608 TSI-----FLKVAFFBELKAVLPKEVEAR-----IAGESGNRAIPRIECCR 651
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      672 TRVLYSFVRBELGVKARGDVFLLQKQEVTTGSNVSRITYEALKSGRLNXLVVKML 725
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      652 SYPLKFFVREGIGT-----ELLTGEKVRSPGECDDKVFAMCEGSIIDPLECL 700

RESULT 7
09M567 PRELIMINARY; PRT; 730 AA.
ID 09M567;
AC 09M567;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN PAL2.
OS Rubus idaeus (Raspberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Rosales; Rosaceae; Rosoideae; Rubus.
OX NCBI_TaxID=32247;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21437948; PubMed=11553751;
RA Kumar A., Ellis B.E.;
RT "The phenylalanine ammonia-lyase gene family in raspberry: structure,
RT expression, and evolution.";
RL Plant Physiol. 127:230-239(2001).
EMBL; AF237955; AAF40224.1; -.
DR HSSP; P21310; 1B8F.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0005559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRPFAM; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 730 AA; 79422 MW; 0B542417B50CE113 CRC64;

Query Match 28.4%; Score 955; DB 10; Length 730;
Best Local Similarity 36.8%; Pred. No. 1.1e-66;
Matches 257; Conservative 107; Mismatches 269; Indels 52; Gaps 19;

QY 40 GSXLPPTXXTQDLYEXXLIADPXTDDXXELDGYSTLGDVVGAAKRGKXRVVXSDERX 99
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 44 GAAEBMTGSHLDEVRWVAE-YRKPVVKGGERTLTISQVAIAINHDSGVVLAESARA 102
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 100 KIDKSYEFLRQQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPFSXDSF 157
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 103 GYKASSDWMDSMNKGTDSYGVTTGFGATSHRRTKOGALQKELIRFLNAGVLANGTES- 161
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

QY 158 XLGGLNSLPLEVVRCAMTIRVNSLTRGSHANRVLVLTALNPLNGHTPIVPIRGITIS 217
D 162 -----AHLTPHSATRAAMLVRIINTLQSGYSGRFLBIAISKFNNHTTPTCLPIRGIT 215
QY 218 ASGDLSPSYIAAATGHPDSKRVKHHEGKEMKAREIALFGLEP--VVGPRGGGL 275
D 216 ASGDLVPSYIAGLITRPNRSKA-VGPKG--ETLNAAEFAQVGISGSGFELQPRGGAL 272
QY 276 VNGTAVASAMATLALDHAKLSLSQALTALTYEAMVAGSGFHPPLHDVTRPHPTQIEV 335
D 273 VNGTAVAGSGLASTLPETNITLALSELISAFIEVVGQKRP-EFTDHLTKLGHHPGQIEA 331
QY 336 ARNRITLLEGSXFVHHEEVKVKDEGLRQDRYPLRTPSPQMLGPIVSDMTHAHAVSL 395
D 332 AATMEHLIDGSSY-VKAAEKLHEDPLQKPKQDRYALRTSPQMLGPIE--VIRSTGSL 388
QY 396 EAG-QSTTNPDLIDVENKXTHGNGFQASAVXNTKTRIALALIGKNTQTLEMLNAG 454
D 389 EREINSVNDPLIDVSHNKALHGGNFQGTPIGVSMNTRLAISIGKLMFAQFSELVNDP 448
QY 455 MNRLPSCLA-AEDPSLYCKGLDIAAAYSELGHLANPTTVVQPRMGNQAVNSLA 513
D 449 YNNGLPSNLISGGRPSLDYGFKAELIAMSYSCELOFLANPTNVQSAEQHNDVNSLG 508
QY 514 LISARTRAEANDVLSLLATHTLCVLAVDLRAMEFEFKQFPYXXXLXKHQFSGXXTX 573
D 509 LISRKTAEAVDILKLMSSFPVALCOALDLHLLENLKSJTKNTVSQLAKRVLTGNG 568
QY 574 XXXXXLXKXVXXKXKLEQINSYDLPRMHDPSXATGTVEKLSXXXXVSLAAVN 633
D 569 ELHSPRECELDLMYVER-EYLPAYIDP-----CSATYPLMQRLQ--VLVEHALTN 618
QY 634 A-----MKVXAEKAS--LTRYRXFXFWXAPSSSPAL--XVLSPTREVLVSF 678
D 619 GENEKASTIFQKIRAFEBELKTLIPKVEESA-RAIYSGNAALPNRIVECRSPLYKF 677
QY 678 VREELGVKARGDVFQKQEVITGSNVSRITYAISKGR 717
D 678 VREELG-----GEPLTGEKVSFGEECDKVFAMCGNI 711

RESULT 8
Q94ENO PRELIMINARY; PRT; 708 AA.
AC Q94ENO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL1.
OS Reimannia glutinosa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Lamiales; incertae sedis; Reimannia.
OX NCBI_TaxID=99300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Keunshan;
RA Yun S.J., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RT "Molecular characterization of phenylalanine ammonia-lyase gene in
RT Reimannia glutinosa."
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF041636; AAK64225.1;
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0016211; F:ammonia lyase activity; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L:aspartase-like.
DR InterPro: IPR001106; Phe/His_NH3-lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR InterPro: IPR006162; Pphanine_S.

```

```

DR Pfam; PF00221; PAL; 1.
DR TIGRfams; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KM Lyase.
SQ SEQUENCE 708 AA; 76835 MW; 1B6AF08B16F1B8FE CRC64;

Query Match 28.4%; Score 954; DB 10; Length 708;
Best Local Similarity 36.7%; Pred. No. 1.3e-66;
Matches 253; Conservative 103; Mismatches 260; Indels 74; Gaps 18;

QY 68 ELDGYSLTGVDVGAAARGKRVYRVYDSEIKTKDKSVEFLRXQJXNSV--YGTPTGFG 125
D 50 KLGESLTIQAQVAA--AAADNVAAYELATAAGVKAASDVMESMNGKTSYGTTTFGA 109
QY 126 SADRTEDALISLQKALEHQLGVLFTSXSFXLGRGLNSLPLEVVRCAMTIRVNSLTR 185
D 110 TSHRRTKQGQALQKELIFFLNAGIFNGTSE-----NHALPHSATRAAMLVRIINTLQ 162
QY 186 GHSAYELVYELTMFLNHGTPVPLRGTSASGGLSPSYIAAATGHPDSKRVKHHE 245
D 163 GTSGRFELTELTLFPLNHNITPCLPIRGITTAGDVLPSYIAGLITRPNRSKA-VGPN 221
QY 246 GXEKIMXAREALATFGLPEPV--LGFKEGLVNGTAVASAMATLALDHAKLSLSQALT 304
D 222 G--EALNNGEASLSLAVGSGFELQKESGLAVNGTAVGSGLASIALYDANILAVSEVTS 279
QY 305 ALTVEAMGHASGFPPLHDVTRPHPTQIEVARNIRTLISGSXFAVHHEEVKVKDEGI 364
D 280 VIFAEVMNGKP-EFTDHLTKLKHHPQIEAAMTEHILDGSAV-VKAAQKHEHTPQK 337
QY 365 LRQDRYPLTPSPQMLGPIVSDMTHAHAVLSAAGQSTTNPDLIDVENKXTHGNGFQASA 424
D 338 PHQDRYALRTSPQMLGPIEIVIRATKRIEEL--NSVNDPLIDVSHNKALHGGNFQGT 396
QY 425 VANTVEKTRIALALIGKNTQTLEMLNAGNRLPSCLA-AEDPSLYCKGLDIAAAA 483
D 397 IGVSMNTRLAIAIGKLMFAQFSELVNDPENNGLPSNLISGGRPSLDYGFKSEIAMS 456
QY 484 YTSSELGLANPTTVVQPRMGNQAVNSLISARTRAEANDVLSLLATHTLCVLAQVD 543
D 457 YCSELQFLANPTNVQSAEQHNDVNSLGLISRKTVEALDILKLMSSFTYLAICOAVD 516
QY 544 LRAME-----FEFKQFPYXXXLXQHF-GXXXTXKXKXKXKXKXKX----- 587
D 517 LHLLENRLSVKNTVSOVAKRTLTMGINGELHPSRCEKDLRVDRVYFAYIDDPQS 576
QY 588 -----LXKRLQCT-----NSVDLEPRMHDPSXATGTVEKLSXXXXVSLAANAKRY 637
D 577 GTFPLMQKLRQVLDVHDALNNGESERN--VSTISFQKIEAF-----VELXAILPKEY 626
QY 638 AXAEKASILTRVXRXFXFWXAPSSSPAL--XVLSPTREVLVSFVREELGVKARGDVFQK 695
D 627 ESARIAL-----ESGNPALGNRITECRSRYPLKFIRELGT-----NYLTG 667
QY 696 KQEVITGSNVSRITYAISKGRINXVLYKML 725
D 668 EKVVSPGEECDKVFALSKGLIVDPLKCL 697

RESULT 9
Q8RUZ3 PRELIMINARY; PRT; 661 AA.
AC Q8RUZ3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
GN PAL1.
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=33349;

```

```

RN [1]
RC SEQUENCE FROM N.A.
EX MEDLINE-21660210; PubMed-11801746;
RA Dvořák V., Sirvio A., Mikkonen M., Savolainen O.,
RT "Low Nucleotide Diversity at the palL Locus in the widely Distributed
  Pinus sylvestris."
RL Mol. Biol. Evol. 19:179-188 (2002).
DR EMBL; AF353967; AAL74317.1; -
DR EMBL; AF353968; AAL74318.1; -
DR EMBL; AF353969; AAL74319.1; -
DR EMBL; AF353970; AAL74320.1; -
DR EMBL; AF353971; AAL74321.1; -
DR EMBL; AF353972; AAL74322.1; -
DR EMBL; AF353973; AAL74323.1; -
DR EMBL; AF353976; AAL74326.1; -
DR EMBL; AF353978; AAL74328.1; -
DR EMBL; AF353979; AAL74329.1; -
DR EMBL; AF353980; AAL74330.1; -
DR EMBL; AF353982; AAL74332.1; -
DR EMBL; AF353983; AAL74333.1; -
DR EMBL; AF353984; AAL74334.1; -
DR EMBL; AF353985; AAL74335.1; -
DR EMBL; AF353986; AAL74336.1; -
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR006948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR TrEMBL; TIGR01226; Phe_am_lyase; 1.
DR TrEMBL; TIGR01226; Phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
KM NON_TER 1 1
FT NON_TER 681 681
SQ SEQUENCE 681 AA; 74365 MW; 5B4521F7978BF804 CRC64;

Query Match 28.1%; Score 944.5; DB 10; Length 681;
Best Local Similarity 43.9%; Pred. No. 7e-66;
Matches 214; Conservative 85; Mismatches 162; Indels 27; Gaps 10;

QY 69 LDGYSLTGADVGAARKGRXVYXDSDEIRKKIDKSVFELRXQLX--NSVGVTTGFGGS 126
DB 56 IEGRSLTISDVAAVARRSQYKVLDAEAKSRVEBSNNVLTQMTKGDTYGVTTGFGAT 115
QY 127 ADRTEDATLQKALLLEHQICGLPTXSXSFYLGRLGSLLELVVVGAMTRVNSLTRG 186
DB 116 SHRTNOGALQKELIRFLNAGV-----LGKCPENVLSDDTTAAVLVNTVTLQG 166
QY 187 HSAVRLVLEALTNFNLHGITPIVPLRGITISAGDSLPLSYIAAATGHPDSKVXVXHEG 246
DB 167 YSGRMNLIETVEKLNAGLTPLRGRITTSAGDVLPLSYIAGLTGRPNRNV--SSDG 225
QY 247 XEKTIXAKKALAFGE--PVILGPKGGLVNGTAVSASMATLADHAMLSLSQALTA 305
DB 226 IE--MSGAEALKKVGKEPPELOPKGGLAVNGTSGAALAIIVCFDANVIALSEVISA 283
QY 306 LTYEAMVGHAGSHPFLHDVTRPHPTQIEVARNIRLTLEGSXFAVH---HBEVKVDD 361
DB 284 MPEEVNNGKEFFDPLTHK--KHHPGOMERAAIMEVYLDGSSYMAAKLHEMPLQKP- 341
QY 362 EGIIRQRIYLRITSPOMLGPVSDMTIAHAVLSLEAGOSTTNPLIDENKXTHHGNFQ 421
DB 342 ----KQRIYLRITSPOMLGPVSDMTIAHAVLSLEAGOSTTNPLIDENKXTHHGNFQ 396
QY 422 ASAVXNMTETLALALIGLNTFLTLEMLNAGMRGLPSCLA--AEDPILSYHCKGLDIA 480
DB 397 GTPIGVSMNLRISALIGLNTFLTLEMLNAGMRGLPSCLA--AEDPILSYHCKGLDIA 456

```

```

QY 481 AAAYTELGHLANPYTHYOPAMEGNOAVNSLALISARTAEANVLSSLATHLYCLO 540
DB 457 MASTSELYLANPVTSIHVQSAEONQDVNSLGLVSAKSAEALIDILKMLSTYALCO 516
QY 541 AYDLRAME 548
DB 517 AYDLRHLE 524

RESULT 10
Q852S1 PRELIMINARY; PRT; 715 AA.
AC Q852S1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN GDCPAL3.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC Daucus.
CX NCBI_TaxID=4039;
RN [1]
RC SEQUENCE FROM N.A.
RA Ozeki Y., Chikigawa Y., Kimura S., Soh H., Maeda K., Pornsiriwong W.,
RA Kato M., Akimoco H., Oyanagi M., Fukuda T., Koda T., Itoh Y.,
RA Yamada A., Ueno H., Nakeda J.,
RT "Putative cis-elements in the promoter region of phenylalanine
  ammonia-lyase gene of carrot involved in anthocyanin biosynthetic
  pathway."
RT Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB089813; ZAC56977.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR006948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TrEMBL; TIGR01226; Phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KM Lyase.
SQ SEQUENCE 715 AA; 77802 MW; 8D067795D57565B4 CRC64;

Query Match 28.1%; Score 944.5; DB 10; Length 715;
Best Local Similarity 36.0%; Pred. No. 7.5e-66;
Matches 251; Conservative 111; Mismatches 250; Indels 85; Gaps 19;

QY 68 ELIDGYSLTGADVGAARKGRXVYXDSDEIRKKIDKSVFELRXQLXSV--YGVTTGFG 124
DB 54 KUGETLITISQVAAISARDSEGVKELSEARAGVKAASDPMWMEENMGTSYGVTTGFG 113
QY 125 GSADTTEDAISLQKALLLEHQICGLPTXSXSFYLGRLG---NSLPLEVVGAMTIVN 181
DB 114 ATSHRTKOGALQKELIRFLNAGI-----FGSGAEAGNNTLPASAVRAAMLVIRIN 164
QY 182 SLTRGSAVRLVLEALTNFNLHGITPIVPLRGITISAGDSLPLSYIAAATGHPDSKVH 241
DB 165 TLQGSIGIRFEILEITFTFLNHNITPCPLRGITTSAGDVLPLSYIAGLTGRPNRNV-- 223
QY 242 VYHEGKXTKAKKALAFGE--PVILGPKGGLVNGTAVSASMATLADHAMLSLSQALTA 299
DB 224 VGPFTV--TLSPDEAFKLAGVEGCFELOPKGGLAVNGTAVGSMAGMVLFEANIIAIVL 281
QY 300 SCALTLTYEAMVGHAGSHPFLHDVTRPHPTQIEVARNIRLTLEGSXFAVHBEVKVX 359
DB 282 AEVMSAIFAEVNVQKP-EFTDHLTHKLNHPQIEAALIMEHILDGSAV--VKAQKHEM 339

```

```

Qy 360 DDEGILRQDRYPLRTSPQWLGPIVSDMIHAHAYLSLEAGOSTTNDPLIDVENKXTHHGN 419
Db 340 DPLQKPKQDRYALRTSPQWLGPIVSDMIHAHAYLSLEAGOSTTNDPLIDVENKXTHHGN 398
Qy 420 FQASAVNTEKTRALALIGKLNFTQTEMNAGNNRGLPSGLA-AEDPSLSYHCKGLD 478
Db 399 FQGPFIQVSDNTRALALIGKLNFTQTEMNAGNNRGLPSGLA-AEDPSLSYHCKGLD 458
Qy 479 IAAAYTSELGHLNPTTHVQPAENGNQAVNSIALISARTRAEANDVSLTLATLHLCV 538
Db 459 IAAAYTSELGHLNPTTHVQPAENGNQAVNSIALISARTRAEANDVSLTLATLHLCV 518
Qy 539 LQAVDLRAMEPEFKQXPRXXXXXLLKHQFGXXXXXELXXXXXKXKX----- 584
Db 519 CQALDLRHLLENLK-----STVKNVTSQVAKRVLNMGVNGELHPSRF 560
Qy 585 -XXIXLRLRQ-----TNSYDLERPMHDAFSXATGVVEXLSXXXXXVSLAAV 632
Db 561 CEKDLRVDBREYFAFYIDPCGATYPLMGKLE-----TLVEHALNNGDKERNLST 613
Qy 633 NAWVAAAEKAKIS--LTRXVXXFWXAPSSSSPAL--XYSPTRVLSFTVEELGVAR 688
Db 614 IFQXIAAFEDLKLALPKVEESA-RAVESGNPAIPNRIRKCRSYPLLYKFAVEELGT--- 669
Qy 689 RGVDFLGKQEVITGSNNVRIYEAIKSGRIXXVLYVKML 725
Db 670 --EYLTGBKVTSPGEFQKVTMTKGEIIDLPLECL 704

```

RESULT 11

Q8AS6 PRELIMINARY; PRT; 681 AA.

```

AC Q8AS6: 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
GN PALI.
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3349;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=21660210; PubMed=11801746;
RA Dvornyk V., Sirvio A., Mikkonen M., Savolainen O.;
RT "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
RT Pinus sylvestris.";
RL Mol. Biol. Evol. 19:179-188 (2002).
DR EMBL: AF353981; AAL74311.1; -.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0016211; F:ammonia lyase activity; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Asparase-like.
DR InterPro: IPR001106; Phe/His NH3lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KW Lyase.
FT NON TER 1 1
FT NON TER 681 681
SQ SEQUENCE 681 AA; 74341 MW; 5BFCC373969FFBD0 CRC64;

```

Query Match 28.0%; Score 943.5; DB 10; Length 681;
 Best Local Similarity 43.9%; Pred. No. 8,4e-66;
 Matches 214; Conservative 85; Mismatches 162; Indels 27; Gaps 10;

```

Qy 69 LDGYSITLGDVGAARKGRXVAVXSDIEIRKKIDKSVFELXQLX--NSVGYTTGFGGS 126
Db 56 IEKSLITSDVAARVRSQVRYKLDAAKARVSSSWVLTQMTKGTDTGVTTGQAT 115
Qy 127 ADRTREDAISLQKALLEHQLCGVLPSTXDSFYLGRGLNSLPLEYVRGAMTIRVNSLTRG 186
Db 116 SHRRTQAGLELQKELLRFNAGV-----LGKCPENVLSEDTTAAALVTRNTLLQG 166
Qy 187 HSAVRLVIEALTNPLNHSITPIVLRGTISASGDSLSLSTIAALIGHPSKHYVHEG 246
Db 167 YSGIEMWDLIEVEKELNGLTPKPLRGITIASGDLVLSYAGLITGRPSRYR-SRDG 225
Qy 247 XEKIKARBAIALFGLF-PVVIQPEKGLVNGFVAVASMATLALHDAHMLSLDSQALTA 305
Db 226 IF--MSGARALKVGLERPFELQPEGLATVNGSVGAALSLYCFDANVALLSVISA 283
Qy 306 LTYEAMVGHAGSFHPLDVTSPHPTQIEVANRITTLIEGSXFVH---HEEVRKYKD 361
Db 284 MFCENVNGKPEFTDPLTHKL-KLHPQMEAAIMBYVDGSSYMKHAKLHNNPLQKP- 341
Qy 362 EGI LRDRYPLRTSPQWLGPIVSDMIHAHAYLSLEAGOSTTNDPLIDVENKXTHHGNFQ 421
Db 342 ---KODRYALRTSPQWLGPIVSDMIHAHAYLSLEAGOSTTNDPLIDVENKXTHHGNFQ 396
Qy 422 ASAVNTEKTRALALIGKLNFTQTEMNAGNNRGLPSGLA-AEDPSLSYHCKGLDIA 480
Db 397 GTFPIGVMNLTLSALIGKLNFTQTEMNAGNNRGLPSGLA-AEDPSLSYHCKGLDIA 456
Qy 481 AAAYTSELGHLNPTTHVQPAENGNQAVNSIALISARTRAEANDVSLTLATLHLCVQ 540
Db 457 MASYTELLYLANPTTHVQPAENGNQAVNSIALISARTRAEANDVSLTLATLHLCVQ 516
Qy 541 AVDLRAME 548
Db 517 AVDLRAME 524

```

RESULT 12

Q9MAX1 PRELIMINARY; PRT; 716 AA.

```

AC Q9MAX1: 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase.
OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Eudicotyledons; core eudicots; asterids;
OC Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vincaceae;
OC Catharanthus.
OX NCBI_TaxID=4058;
RN [1]
RC SEQUENCE FROM N.A.
RX Kiyoca S., Sakano K., Kim U.;
RT "cDNA of phenylalanine ammonia lyase from Catharanthus roseus.";
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB042520; BA95629.1; -.
DR HSP: P21317; 1B8F.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0016211; F:ammonia lyase activity; IEA.
DR GO: GO:0016841; F:ammonia-lyase activity; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0009058; P:biosynthesis; IEA.
DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Asparase-like.
DR InterPro: IPR001106; Phe/His NH3lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 716 AA; 78241 MW; 1F16D4DB8683B0B CRC64;

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:45:43 ; Search time 89.3693 Seconds
(without alignments)
2295.302 Million cell updates/sec

Title: US-09-939-408a-21
Perfect score: 3365
Sequence: 1 NAPSIDSATISXANGXNGX.....RIYEAIKSGRINXVLYKMLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_290a04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3223	95.8	726	5	ABB07694 Rhodotoru
2	3223	95.8	726	5	AAB27942 Phenylala
3	3021	89.8	716	1	AAP80513 L-phenyla
4	3021	89.8	716	5	ABB07693 R. torulo
5	3021	89.8	716	5	AAB27941 Rhodotoru
6	3021	89.8	716	5	AAB20653 Rhodotoru
7	3021	89.8	716	5	AAB16389 Rhodotoru
8	3021	89.8	716	6	ABG71948 R. glutin
9	3021	89.8	716	7	ADE36054 Rhodospo
10	3019	89.7	716	5	AAB20654 Rhodotoru
11	3019	89.7	716	5	AAB16390 Rhodotoru
12	3019	89.7	716	6	ABG71955 R. glutin
13	3019	89.7	716	6	ADE36056 Rhodospo
14	3018	89.7	716	1	AAP81099 Sequence
15	3018	89.7	716	1	ADE36076 Rhodospo
16	3015	89.6	716	1	AAP83141 Sequence
17	3015	89.6	716	5	ADE36074 Rhodospo
18	3015	89.6	716	5	AAB20663 Rhodotoru
19	3014	89.6	716	7	ABG71954 R. glutin
20	3014	89.6	716	7	ADE36068 Rhodospo
21	3013	89.5	716	5	AAB20667 Rhodospo
22	3013	89.5	716	7	ADE36060 Rhodospo
23	3012	89.5	716	5	AAB20665 Rhodospo
24	3012	89.5	716	5	AAB20711 R. glutin
25	3012	89.5	716	5	AAB20711 R. glutin

26	3012	89.5	716	5	AAB20660 Rhodotoru
27	3012	89.5	716	6	ABG71951 R. glutin
28	3012	89.5	716	7	ADE36062 Rhodospo
29	3011	89.5	716	7	ADE36066 Rhodospo
30	3010	89.5	716	5	AAB20662 Rhodotoru
31	3010	89.5	716	5	AAB20661 Rhodotoru
32	3010	89.5	716	5	AAB20666 R. glutin
33	3010	89.5	716	6	ABG71952 R. glutin
34	3010	89.5	716	6	ABG71953 L-phenyla
35	3007	89.4	716	1	AAP95783 Rhodotoru
36	3006	89.3	716	5	AAB20659 Rhodotoru
37	3006	89.3	716	6	ABG71950 R. glutin
38	3006	89.3	716	7	ADE36070 Rhodospo
39	3005	89.3	716	5	AAB20664 Rhodospo
40	3005	89.3	716	5	ADE36058 Rhodospo
41	3005	89.3	716	7	ADE36072 Rhodospo
42	2997	89.1	716	7	ADE36064 Rhodospo
43	2978	88.5	716	5	AAB20658 Rhodotoru
44	2978	88.5	716	6	ABG71949 R. glutin
45	2971	88.3	686	5	AAB20657 Rhodotoru

ALIGNMENTS

RESULT 1
ABB07694 standard; protein; 726 AA.
XX
AC ABB07694;
XX
AC 29-ANG-2003 (revised)
DT 07-AUG-2003 (revised)
DT 10-JUN-2002 (first entry)
XX
XX Rhodotorula PAL consensus polypeptide sequence.
DE
XX
XX PAL, yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KW consensus; enzyme.
XX
XX Rhodotorula graminis.
OS Rhodotorula mucilaginosa.
OS Rhodotorula mucilaginosa.
OS Rhodospiridium toruloides.
OS
XX
XX
XX Key Location/Qualifiers
FH MISC-difference 1..726
FT /note= "Xaa is any amino acid; "Xaa" indicates no
FT consensus at that position"
XX
XX W0200208402-A2.
XX
XX 31-JAN-2002.
XX
XX 24-JUL-2001; 2001MO-US023270.
XX
XX 24-JUL-2000; 2000US-00624693.
XX
XX (PCBU-) PCBU SERVICES INC.
XX
XX Yoshida RK, Kootstra AB;
XX WPI; 2002-268973/31.
XX N-PSDB; ABA95244.
XX
XX Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
XX treating mammal having disease or disorder from phenylketonuria, cancer,
XX human immunodeficiency virus infection and human cytomegalovirus
XX infection.
XX
XX Claim 10; Fig 2; 135pp; English.
XX
XX

Db 181 NSLTGSHAVLVLEALTNFLNHTGIVPLRGTHISAGSLPSLYIAAATGHPDSKV 240
 QY 241 HXHEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 300
 Db 241 HXHEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 300
 QY 301 QALTAITYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
 Db 301 QALTAITYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
 QY 361 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 420
 Db 361 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 420
 QY 421 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 480
 Db 421 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 480
 QY 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 540
 Db 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 540
 QY 541 AVDLAMEEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXXKXRLBQTNSTYDL 600
 Db 541 AVDLAMEEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXXKXRLBQTNSTYDL 600
 QY 601 EPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 660
 Db 601 EPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 660
 QY 661 SSPALXYLSPTRTVLYSPFREELGVKARGDVFLEKQVTTGNSVSRVYEAIRKGRINXY 720
 Db 661 SSPALXYLSPTRTVLYSPFREELGVKARGDVFLEKQVTTGNSVSRVYEAIRKGRINXY 720
 QY 721 LVKMLA 726
 Db 721 LVKMLA 726
 Db 721 LVKMLA 726
 RESULT 3
 AAB0513
 ID AAB0513 standard; protein; 716 AA.
 AC AAB0513;
 XX 25-MAR-2003 (revised)
 DT 12-NOV-1990 (first entry)
 XX L-phenylalanine ammonialyase.
 DE L-phenylalanine ammonialyase.
 XX L-phenylalanine ammonialyase; PAL; foreign gene expression;
 KM culture temperature; expression regulation.
 OS Rhodospiridium toruloides.
 XX EP279665-A.
 PN 24-AUG-1988.
 PD 18-FEB-1988; 88EP-00301356.
 PF 19-FEB-1987; 87GP-00034397.
 PR 18-JUN-1987; 87JP-00152359.
 XX (MITK) MITSUI TOATSU CHEM. INC.
 XX Fukuhiara N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y,
 DR WPI; 1988-236895/34.
 DR N-PSDB; AAN81116.
 PT Regulation of expression of foreign gene in Escherichia coli - by

PT maintaining temp. of culture at 40 deg. C or more to suppress expression.
 XX Claim 3; Page 15-19; 36pp; English.
 CC The PAL gene is expressed in E.coli carrying a recombinant plasmid.
 CC Expression is regulated by maintaining the temperature at at least 40
 CC deg. C to suppress expression. Pal is produced in high concentrations.
 CC See also EP-279664; regulation by sugar component as C source at 0.3% or
 CC more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
 CC 2003 to correct PI field.)
 XX
 SQ Sequence 716 AA;
 Query Match 89.8%; Score 3021; DB 1; Length 716;
 Best Local Similarity 84.0%; Pred. No. 4.5e-300;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;
 QY 1 MAPSDISTATXANGXNKGHAAXXASXXXXXXXASXKPTXXTOLDIVEXXAD 60
 Db 1 MAPSDISTATXANGXNKGHAAXXASXXXXXXXASXKPTXXTOLDIVEXXAD 60
 QY 61 PXTDXEILDGYSILGVDVGAARKGXVYXSDERKXKIDKSVFLRXQLXNSVYGT 120
 Db 61 PXTDXEILDGYSILGVDVGAARKGXVYXSDERKXKIDKSVFLRXQLXNSVYGT 120
 QY 121 TGFSSADTRFEDALSLOKALLEHOLGVLPSTXSPXLRGLENSTPLEVYGAMTIRV 180
 Db 121 TGFSSADTRFEDALSLOKALLEHOLGVLPSTXSPXLRGLENSTPLEVYGAMTIRV 180
 QY 114 TGFSSADTRFEDALSLOKALLEHOLGVLPSTXSPXLRGLENSTPLEVYGAMTIRV 173
 Db 114 TGFSSADTRFEDALSLOKALLEHOLGVLPSTXSPXLRGLENSTPLEVYGAMTIRV 173
 QY 181 NSLTGSHAVLVLEALTNFLNHTGIVPLRGTHISAGSLPSLYIAAATGHPDSKV 240
 Db 181 NSLTGSHAVLVLEALTNFLNHTGIVPLRGTHISAGSLPSLYIAAATGHPDSKV 240
 QY 174 NSLTGSHAVLVLEALTNFLNHTGIVPLRGTHISAGSLPSLYIAAATGHPDSKV 233
 Db 174 NSLTGSHAVLVLEALTNFLNHTGIVPLRGTHISAGSLPSLYIAAATGHPDSKV 233
 QY 241 HXHEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 300
 Db 241 HXHEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 300
 QY 234 HVHDEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 293
 Db 234 HVHDEGKEXIKKARBAIALFGLPEVYVGPXKGLGIVNGTAASAMATLALHDAMLSLS 293
 QY 301 QALTAITYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
 Db 301 QALTAITYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
 QY 294 QSLTMTVEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 353
 Db 294 QSLTMTVEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 353
 QY 361 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 420
 Db 361 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 420
 QY 354 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 413
 Db 354 DEGLIRQDRYPLRTSPQWLGPLVSDMITHAAVLSLEAGOSTTNDPLIDVENKXTHHGNT 413
 QY 421 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 480
 Db 421 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 480
 QY 414 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 473
 Db 414 QASAVNTEKTRIALALIGKLNFTQITEMNAGNRLPSCLAEDPSLSYHCKGIDIA 473
 QY 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 540
 Db 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 540
 QY 474 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 533
 Db 474 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISARTEANDVLSLLATHTLVCVLQ 533
 QY 541 AVDLAMEEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 600
 Db 541 AVDLAMEEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 600
 QY 534 AIDLRALIEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 593
 Db 534 AIDLRALIEFEPKQRPXKXXLLXGHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 593
 QY 601 EPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 660
 Db 601 EPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 660
 QY 594 VPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 650
 Db 594 VPRMHDASFATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRXXVXXFWXAPSS 650
 QY 661 SSPALXYLSPTRTVLYSPFREELGVKARGDVFLEKQVTTGNSVSRVYEAIRKGRINXY 720
 Db 661 SSPALXYLSPTRTVLYSPFREELGVKARGDVFLEKQVTTGNSVSRVYEAIRKGRINXY 720
 QY 721 LVKMLA 726
 Db 721 LVKMLA 726
 Db 711 LVKMLA 716
 RESULT 4
 AAB07693
 ID AAB07693 standard; protein; 716 AA.

XX ABB07693;
AC
XX 07-AUG-2003 (revised)
DT
DT 10-JUN-2002 (first entry)
XX
XX R. toruloides PAL polypeptide.
DE
XX PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
KW cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
XX Rhodospiridium toruloides.
OS
XX WC020208402-A2.
FN
XX 31-JAN-2002.
PD
XX 24-JUL-2001; 2001MO-US023270.
PE
XX 24-JUL-2000; 2000US-00624693.
PR
XX (PCBU-) PCBU SERVICES INC.
PA
XX Yoshida RK, Kootstra AB;
PI
XX WPI; 2002-268973/31.
DR
DR N-PSDB; ABA95243.
XX
XX Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT treating mammal having disease or disorder from phenylketonuria, cancer,
PT human immunodeficiency virus infection and human cytomegalovirus
PT infection.
XX
XX Example 3; Fig 1; 135pp; English.
PS
XX The invention relates to an isolated and purified yeast phenylalanine
CC ammonia-lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC host cell comprising the polynucleotide is useful for the production of L
CC -PAL, by adding the polynucleotide to the host cell to a composition
CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC improving the production of PAL, its analog or another optically active
CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC are useful for treating a mammal having a disease, disorder or condition
CC selected from phenylketonuria, cancer, human immunodeficiency virus
CC infection and human cytomegalovirus infection. The present sequence
CC represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
CC correct OS field.)
XX
XX
SQ Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;
Best Local Similarity 84.0%; Pred. No. 4, 5e-300; Indels 10; Gaps 3;
Matches 610; Conservative 25; Mismatches 81;

QY 1 MAPSLDLSIATXNAGXNGXHAAXXAXXXXXXXXAXAGXLPPTXXXTQLDIVEXXLIAD 60
DB 1 MAPSLDLSIATXNAGXNGXHAAXXAXXXXXXXXAXAGXLPPTXXXTQLDIVEXXLIAD 60
QY 61 PXTDDXKXELDGYSLTIGDVVGAARKGRXRVVDSDEIKXIDKXSEFIRKXLDXNXYGVY 120
DB 55 P-TDSTLELDGYSLNIGDVVGAARKGRPVVRSDEIRSKIDKXSEFIRKXLDXNXYGVY 113
QY 121 TGFSGSADRTEDAIISQKALDLEHOLCVLPSTXSPFLGSGLENSLPLEVRGAMTIRV 180
DB 114 TGFSGSADRTEDAIISQKALDLEHOLCVLPSTXSPFLGSGLENSLPLEVRGAMTIRV 173
QY 181 NSLITRHSASVRLVYLTALNFTNPGITPIVPLRGITISASGULSPUSYIAAAITGHPDSKV 240
DB 174 NSLITRHSASVRLVYLTALNFTNPGITPIVPLRGITISASGULSPUSYIAAAITGHPDSKV 233
QY 241 HXVHEGEXEKIMVAREALFGLPEVYLGPKSGELGVNGTAVASASMAATLALHDAHMLSL 300
DB 234 HXVHEGEXEKILVAREMALFNLEPVYLGPKSGELGVNGTAVASASMAATLALHDAHMLSL 293

QY 301 QALTALTYEAMYGHAGSFHPLHDYTRPHPTQIOIEAANIRTLLEGSGYFAVHHEEVKVD 360
DB 294 QSLTAMTYEAMYGHAGSFHPLHDYTRPHPTQIOIEAANIRTLLEGSGYFAVHHEEVKVD 353
QY 361 DEGIILRODRYPLRTSPQWLGPLVSDMLHAHVALSLLEAGQSTTDNPLIDVENKXTHGNGF 420
DB 354 DEGIILRODRYPLRTSPQWLGPLVSDMLHAHVALSLLEAGQSTTDNPLIDVENKXTHGNGF 413
QY 421 QASAYXNMTKRLALALIGKLNFTQLTENLNAQMNGLBSCLAEPSPSYHCKGIDIA 480
DB 414 QAAAVANTMEXTRGLAIGLNFTQLTENLNAQMNGLBSCLAEPSPSYHCKGIDIA 473
QY 481 AAAYTSELGHLANPTTVQPAEMGNQAVNSIALISARRTEANDVLSLLATHLVCLQ 540
DB 474 AAAYTSELGHLANPTTVQPAEMGNQAVNSIALISARRTESNDVLSLLATHLVCLQ 533
QY 541 AYDLRAMEFEFEKQXPXXXXLXQHFQXXXTXXXXXXELXXKXKXKXRLQNTNSYDL 600
DB 534 AIDLRAIFEFEKQGPALVSLIDQHFQSAMTGSNLRDEVEKXNKTALRLEQNTNSYDL 593
QY 601 EPRMHDASFXYATGTVEXTLSXXXXXVSLAANVAMKVAAXEKAISLTRVXXFXWAPSS 660
DB 594 VPRMHDASFXYATGTVEXTLSXXXXXVSLAANVAMKVAAXEKAISLTRVXXFXWAPSS 650
QY 661 SSPALXYLSPTRTVLVYSFREELGVKARQGVFLGKQEVTTIGSNVSRIVYALIKGRINXY 720
DB 651 SSPALXYLSPTRTVLVYSFREELGVKARQGVFLGKQEVTTIGSNVSRIVYALIKGRINXY 710
QY 721 LVKMLA 726
DB 711 LVKMLA 716

RESULT 5

AAE27941
ID AAE27941 standard; protein; 716 AA.

AC AAE27941;
XX
XX 06-AUG-2003 (revised)
DT 27-DEC-2002 (first entry)
XX

DE Rhodotorula toruloides PAL protein.

XX Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; HCMV;
KW human cytomegalovirus infection; cytosolic; virucide.
XX

OS Rhodospiridium toruloides.

XX US2002102712-A1.

PN 01-AUG-2002.

PD 24-AUG-2001; 2001US-00939408.

PE 24-JUL-2000; 2000US-00624693.

PR 24-JUL-2001; 2001MO-US023270.

XX (PCBU-) PCBU SERVICES INC.

XX Yoshida RK, Kootstra AB;

XX WPI; 2002-690616/74.

XX N-PSDB; AAD45815.

XX Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.

XX Example 3; Page 43-45; 74pp; English.

CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC They are useful for treating mammals having diseases, disorders or
 CC conditions that would benefit from treatment with PAL proteins such as
 CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
 CC human cytomegalovirus (hcmv) infection. Sequences of the invention are
 CC also used in gene therapy. The present sequence is R. toruloides PAL
 CC protein. (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;

Best Local Similarity 84.0%; Pred. No. 4.5e-300;

Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MADSLSIATSKXNGXNGXHAAXXASXXXXXXXAXGSLPTTXXXTOLDIVEXXAD 60
 DB 1 MADSLSISHSFANGVSAKQAVNGAS-----TNLAAGSHLPTTQVTDIVEXMLAA 54
 QY 61 PXTDDXXEIDGYSILTLGDPVGAARKGRXRVYXDSDEIRKXIDKVEFLRXQLXNSVYGV 120
 DB 55 P-TDSTLELDGYSILNLGDPVSAARKGRPVRYKDSDEIRSKIDSVEFLRSQLSNSVYGV 113
 QY 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENLPLEVVGAMTIRV 180
 DB 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENLPLEVVGAMTIRV 173
 QY 181 NSLTRGHSAVRLVYLFALNPLNHTGTPVPLRGTTASAGDISPLSYIAAIIHGPDGSKV 240
 DB 174 NSLTRGHSAVRLVLEALNPLNHTGTPVPLRGTTASAGDISPLSYIAAIIHGPDGSKV 233
 QY 241 HVHHEGEXKIMKAREALFGLPEVYLGPKEGLGVNGTAVASASMATLADAHMLSLLS 300
 DB 234 HVHHEGEXKILVAREMALFNLPEVYLGPKEGLGVNGTAVASASMATLADAHMLSLLS 293
 QY 301 QALTALTEAMVGHAGSEHPLHDVTRPHPTQIEVARNIRTLLEGSPAVHHEEVAKVD 360
 DB 294 QSLTALTEAMVGHAGSEHPLHDVTRPHPTQIEVAGNIRKLESGRPVHHEEVAKVD 353
 QY 361 DEGLIQRDSYPLRTSPQWLGPIVSDMHNHAYLSLAGOSTTNDPLDVENKTHHGNF 420
 DB 354 DEGLIQRDSYPLRTSPQWLGPIVSDMHNHAYLSLAGOSTTNDPLDVENKTHHGNF 413
 QY 421 QASAVNTEKTRIALALIGKINFTQITEMNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
 DB 414 QASAVNTEKTRIALALIGKINFTQITEMNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
 QY 481 AAAYTSELHLNPTTTHVQPAEMNGQAVNSIALISARTIADANDVLSLLATHLVCYLQ 540
 DB 474 AAAYTSELHLNPTTTHVQPAEMNGQAVNSIALISARTIADANDVLSLLATHLVCYLQ 533
 QY 541 AVDLRAMEEFKQEPFXKXXLLXQHPGKXXYXXXXXXELKXXYKXILKRLQGTNSYDL 600
 DB 534 AIDLRAIEFEFKQEPFAPVSLIDQHPGSAMTGSMLRELVENKNTAKRLQGTNSYDL 593
 QY 601 EPRMHDASEXAFTVYEXLSSXXXXVSLAAVNANKVAAKKAISLTRYXXFXFWAPSS 660
 DB 594 VPRMHDASEFAAGTVEVLS--TSLSLAAVNANKVAAASALSLTFQVETFWMAST 650
 QY 661 SSPALXYLSPTRTVLYSFVREELGVKARSGDVEFLGKQVTTGNSVSRVYEIRKGRINXV 720
 DB 651 SSPALXYLSPTRTVLYSFVREELGVKARSGDVEFLGKQVTTGNSVSKLYEAIKSGRINNV 710
 QY 721 LVEMLA 726
 DB 711 LKMLA 716

RESULT 6
 AAE20653
 ID AAE20653 standard; protein; 716 AA.

AC AAE20653;
 XX
 XX 01-JUL-2002 (first entry)
 XX
 DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL).
 KW Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 KW LCP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst.
 XX
 OS Rhodotorula glutinis.
 XX
 PN WO200210407-A1.
 PD 07-FEB-2002.
 PF 23-JAN-2001; 2001WO-US002099.
 PR 27-JUL-2000; 2000US-00627216.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Gatenby AA, Sariaslani FS, Tang X, Qi WW, Vannelli T;
 DR WPI; 2002-280635/32.
 XX N-PSDB; AAD33075.
 PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.
 XX
 PS Example 6; Page 92-95; 139pp; English.
 XX
 CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
 CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used
 CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC PAL protein
 CC
 XX
 SQ Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;
 Best Local Similarity 84.0%; Pred. No. 4.5e-300;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MADSLSIATSKXNGXNGXHAAXXASXXXXXXXAXGSLPTTXXXTOLDIVEXXAD 60
 DB 1 MADSLSISHSFANGVSAKQAVNGAS-----TNLAAGSHLPTTQVTDIVEXMLAA 54
 QY 61 PXTDDXXEIDGYSILTLGDPVGAARKGRXRVYXDSDEIRKXIDKVEFLRXQLXNSVYGV 120
 DB 55 P-TDSTLELDGYSILNLGDPVSAARKGRPVRYKDSDEIRSKIDSVEFLRSQLSNSVYGV 113
 QY 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENLPLEVVGAMTIRV 180
 DB 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENLPLEVVGAMTIRV 173

```

QY 181 NSLTRGHSAYRLVLEALTNPLNHGTTPIVPLRGTTISASGDLSPSTYATAAITGHPSKV 240
D 174 NSLTRGHSAYRLVLEALTNPLNHGTTPIVPLRGTTISASGDLSPSTYATAAISHGPSKV 233
QY 241 HVXHEGXEKIMKAREALFGLPEVVLGPKGLGVNGTAVSASMATLALDHAMLSLS 300
D 234 HVXHEGXEKILYAREAMALFNLPEVVLGPKGLGVNGTAVSASMATLALDHAMLSLS 293
QY 301 QALTALVEAMNGHAGSFPHLDVTRPHPTQIEVARNRITLLESGFAVHHEEVKVD 360
D 294 QSLTAMTVEAMVGHAGSFPHLDVTRPHPTQIEVAGNRITLLESGFAVHHEEVKVD 353
QY 361 DEGITRQDRYPLRTSPQWLGPIVSDMIHAHVALSLEAGOSTTDNPLIDVENKXTHGNGF 420
D 354 DEGITRQDRYPLRTSPQWLGPIVSDMIHAHVALTEAGOSTTDNPLIDVENKXTHGNGF 413
QY 421 QASAVXNTEKTRALALIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 480
D 414 QAAAVANTMEKTRGLAOLGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARTEANDVLSLLATLHYCVLQ 540
D 474 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARTEANDVLSLLATLHYCVLQ 533
QY 541 AVDLRAMEFEFKQKQFPKXXXLLXQHPGXXXTXXXXXLXXKXKXKXKXLEQTSYDL 600
D 534 AIDLRATIEFEFKQKQFPKXXXLLXQHPGXXXTXXXXXLXXKXKXKXKXLEQTSYDL 593
QY 601 EPRWHDFAFXATGTVEVEXLSXXXXXVSLAANAMVAAAEKALSTEXVXXFXPAASS 660
D 594 VPRWHDFAFXATGTVEVEXLSXXXXXVSLAANAMVAAAEKALSTEXVXXFXPAASS 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARAGDVLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
D 651 SSPALXYLSPTRTVLVSFVREELGVKARAGDVLGKQEVTTIGSNVSRITYEAIKSGRINXV 710
QY 721 LVKMLA 726
D 711 LVKMLA 716

RESULT 7
AAE16389
ID AAE16389 standard; protein; 716 AA.
XX
AC AAE16389;
DT
09-APR-2002 (first entry)
XX
D 09-APR-2002 (first entry)
XX
D Rhodocorula glutinis wildtype phenylalanine ammonia-lyase (PAL).
XX
KW Phenylalanine ammonia-lyase; PAL; enzyme; citramate; food packing; PHCA;
KW para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW aerospace application; chemical; red yeast.
XX
OS Rhodocorula glutinis.
XX
W020011071-A2.
XX
15-FEB-2001.
XX
03-AUG-2000; 2000WO-US021156.
XX
06-AUG-1999; 99US-01477192.
XX
(DDPO) DU PONT DE NEMOURS & CO E I.
XX
Tang X, Vannelli TM, Qi WW, Sarisiani S, Gatenby AA;
XX
WPI; 2002-121549/16.
XX
DR N-PSDB; AAD26923.
XX

```

```

PT Producing para-hydroxycinnamic acid for producing liquid crystal polymer;
PT comprises converting citramate to PHCA, glucose to PHCA by phenyl ammonia
PT -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT activity.
XX
PS Claim 8; Page 65-67; 75pp; English.
XX
CC The invention relates to a method for producing para-hydroxycinnamic acid
CC (PHCA). The method comprises converting citramate to PHCA by converting
CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC route or by generating a new biocatalyst possessing enhanced tyrosine
CC ammonia-lyase (PAL) activity. The method is useful for the biological
CC production of PHCA which is useful as a monomer for the production of
CC liquid crystal polymers (LCP), where the LCP is used in electronic
CC connectors, telecommunications and aerospace applications. LCP is also
CC useful in medical devices, as well as chemical, and food packing
CC applications due to its resistance to sterilizing radiation. The present
CC sequence is red yeast, Rhodocorula glutinis (Rhodospiridium toruloides)
CC wildtype PAL
XX
SQ Sequence 716 AA;
XX
Query Match 89.8%; Score 3021; DB 5; Length 716;
Best Local Similarity 84.0%; Pred. No. 4,56-300;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;
QY 1 NAPS.LDSIATSYXNGXNGXGHAAXXAXXXXXXXXAXGSKLPTTXXTQLDIYEXXLD 60
D 1 NAPS.LDSIHSFANQVAVASAKQAVNGAS-----TNLVAGSHLPTTQVTDIVKMLAA 54
QY 61 EYTDXXELDGYSLTGLGVVGAARKGRVYRSDDEIRKXDKVEFLRQXNLSVYGV 120
D 55 P-TDSTLELDGYSLTGLGVVGAARKGRVYRSDDEIRKXDKVEFLRQXNLSVYGV 113
QY 121 TGFSGSADRTEDALSLQKALLEHQLCGVLPTSDXDFLXGLGELNSLPLEVVGAMTIRV 180
D 114 TGFSGSADRTEDALSLQKALLEHQLCGVLPTSDXDFLXGLGELNSLPLEVVGAMTIRV 173
QY 181 NSLTRGHSAYRLVLEALTNPLNHGTTPIVPLRGTTISASGDLSPSTYATAAISHGPSKV 240
D 174 NSLTRGHSAYRLVLEALTNPLNHGTTPIVPLRGTTISASGDLSPSTYATAAISHGPSKV 233
QY 241 HVXHEGXEKIMKAREALFGLPEVVLGPKGLGVNGTAVSASMATLALDHAMLSLS 300
D 234 HVXHEGXEKILYAREAMALFNLPEVVLGPKGLGVNGTAVSASMATLALDHAMLSLS 293
QY 301 QALTALVEAMNGHAGSFPHLDVTRPHPTQIEVARNRITLLESGFAVHHEEVKVD 360
D 294 QSLTAMTVEAMVGHAGSFPHLDVTRPHPTQIEVAGNRITLLESGFAVHHEEVKVD 353
QY 361 DEGITRQDRYPLRTSPQWLGPIVSDMIHAHVALSLEAGOSTTDNPLIDVENKXTHGNGF 420
D 354 DEGITRQDRYPLRTSPQWLGPIVSDMIHAHVALTEAGOSTTDNPLIDVENKXTHGNGF 413
QY 421 QASAVXNTEKTRALALIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 480
D 414 QAAAVANTMEKTRGLAOLGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARTEANDVLSLLATLHYCVLQ 540
D 474 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARTEANDVLSLLATLHYCVLQ 533
QY 541 AVDLRAMEFEFKQKQFPKXXXLLXQHPGXXXTXXXXXLXXKXKXKXKXLEQTSYDL 600
D 534 AIDLRATIEFEFKQKQFPKXXXLLXQHPGXXXTXXXXXLXXKXKXKXKXLEQTSYDL 593
QY 601 EPRWHDFAFXATGTVEVEXLSXXXXXVSLAANAMVAAAEKALSTEXVXXFXPAASS 660
D 594 VPRWHDFAFXATGTVEVEXLSXXXXXVSLAANAMVAAAEKALSTEXVXXFXPAASS 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARAGDVLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
D 651 SSPALXYLSPTRTVLVSFVREELGVKARAGDVLGKQEVTTIGSNVSRITYEAIKSGRINXV 710

```

QY 721 LVKMLA 726
 Db 711 LKMLA 716

RESULT 8
 ID ABG71948 standard; protein; 716 AA.
 ABG71948

XX ABG71948;

DT 27-JAN-2003 (first entry)

DE R. glutinis wild-type phenylalanine ammonia-lyase, PAL.

XX Enzyme; phenylalanine ammonia-lyase, PAL; phenylalanine hydroxylase, PAH;
 KM para-hydroxycinnamic acid; tyrosine ammonia-lyase, TAL; tyrosine;
 XX cinnamic acid; PHCA; liquid crystal polymer.

OS Rhodocetula glutinis.

XX WO200290523-A2.

XX 14-NOV-2002.

PF 03-MAY-2002; 2002WO-US018551.

XX 04-MAY-2001; 2001US-0288701P.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX Qi WW, Sarisiani FS, Tang X;

DR WPI; 2003-058928/05.

XX N-PSDB; AB856530.

PT Novel recombinant host for producing para-hydroxycinnamic acid and
 PT tyrosine, comprises gene encoding tyrosine ammonia-lyase activity and
 PT gene encoding phenylalanine hydroxylase activity.

PS Claim 5; Page 48-50; 69pp; English.

XX The invention relates to a recombinant host comprising at least one gene
 CC encoding a tyrosine ammonia-lyase (TAL) activity, and at least one gene
 CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
 CC providing the recombinant host; and (2) growing the recombinant organism
 CC in the presence of a fermentable carbon substrate and thus producing para-
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
 CC providing a recombinant organism comprising at least one gene encoding a
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant
 CC organism in the presence of a fermentable carbon substrate to produce
 CC tyrosine. The recombinant host cell is used for producing para-
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
 CC monomer for the production of liquid crystal polymer. The present
 CC sequence is R. glutinis wild-type phenylalanine ammonia-lyase, PAL, an
 CC enzyme which catalyses the conversion of phenylalanine into cinnamic acid

XX Sequence 716 AA.

Query Match 89.8%; Score 3021; DB 6; Length 716;
 Best Local Similarity 84.0%; Pred. No. 4.5e-300;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDISATSKXNGXNGXGAAXXASXXXXXXAXXGSLPTTXXTQDLDIVEXXAD 60
 Db 1 MAPSLDISATSKXNGXNGXGAAXXASXXXXXXAXXGSLPTTXXTQDLDIVEXXAD 60
 QY 61 PXTDDXXEIDGYSITLIGDVVGAARKGRVYVXDSDEIRKKIDKSVFFLRXQLANSVYGT 120
 Db 55 P-TDSTLELDGYSINLIGDVVGAARKGRVYVXDSDEIRKKIDKSVFFLRXQLANSVYGT 113

QY 121 TGFGGSAADRTEDALISLOKALLEHQLGCVLPTXKDSFXLGRGLNSLPLEVVGAMTIRV 180
 Db 114 TGFGGSAADRTEDALISLOKALLEHQLGCVLPSFDSFRLGRGLNSLPLEVVGAMTIRV 173
 QY 181 NSLTRGHSVAVRLVLEALTNFLNNGITPIVPLRGTISASGDLSPLSYIAAITGHPDSKV 240
 Db 174 NSLTRGHSVAVRLVLEALTNFLNNGITPIVPLRGTISASGDLSPLSYIAAISGHPDSKV 233
 QY 241 HXHEGKXIKKARALVRLGLEPVVLPKKEGLGVNNGTAVSASMATLALHDAHMSLLS 300
 Db 234 HXHEGKXIKKARALVRLGLEPVVLPKKEGLGVNNGTAVSASMATLALHDAHMSLLS 293
 QY 301 QALTALTEAMVGHAGSFHPPLHDVTRPHPTQIEVANIRTLLEGSAFVAHHEBEVVKD 360
 Db 294 QSLTAMTIEAMVGHAGSFHPPLHDVTRPHPTQIEVANIRTLLEGSAFVAHHEBEVVKD 353
 QY 361 DEGIIRQDRYPLRTSPQWLGPLVSDMTAAVLSLEKQSTTDNPLIDVENKTHHGNF 420
 Db 354 DEGIIRQDRYPLRTSPQWLGPLVSDMTAAVLSLEKQSTTDNPLIDVENKTHHGNF 413
 QY 421 QASAVXNTMEKTRRLALIGKLNFTQLEMLNAGMNRGLPSCLAAPPSLSYHCKGLDIA 480
 Db 414 QASAVXNTMEKTRRLALIGKLNFTQLEMLNAGMNRGLPSCLAAPPSLSYHCKGLDIA 473
 QY 481 AAAYTSELGHLANPYTHVQPAEWGNQAVNSLALISARPTAEANDVLSLIATHLYCYLQ 540
 Db 474 AAAYTSELGHLANPYTHVQPAEWGNQAVNSLALISARPTESNDVLSLIATHLYCYLQ 533
 QY 541 AYDLRAMEEFKKQCPKXXXXLXKHQFGKXTXXXXXELKKYKXKXLEQTNSTYL 600
 Db 534 AIDLRALIEFEPKQGPALIVSLIDQFESAMTGSNLRBELVEKNKTLAKLEQTNSTYL 593
 QY 601 EPRWDAFESXATGTVVEXLSXXXXXVSLAAVNAKVAAXAKAISLTXVXXKFWXAPSS 660
 Db 594 EPRWDAFESXATGTVVEXLSXXXXXVSLAAVNAKVAAXAKAISLTXVXXKFWXAPSS 650
 QY 661 SSPALXYSPPRRVYISFVREELGYKARGDVFLLGKQVTTIGSVSRYYEIKSGRINXV 720
 Db 651 SSPALXYSPPRRVYISFVREELGYKARGDVFLLGKQVTTIGSVSRYYEIKSGRINXV 710

RESULT 9

AD836054
 ID ADE36054 standard; protein; 716 AA.

XX ADE36054;

XX 29-JAN-2004 (first entry)

XX Rhodosporidium glutinis PAL protein SEQ ID NO:2.

XX recombination; nucleic acid template; recombinant polypeptide;

XX altered property; combined gene.

XX Rhodosporidium.

XX WO2003072743-A2.

XX 04-SEP-2003.

XX 26-FEB-2003; 2003WO-US005708.

XX 26-FEB-2002; 2002US-0360279P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Milano U, Tang X;

DR WPI: 2003-721766/68.
 DR N-PSDB; ADE36053.
 PT Recombination of nucleic acid templates, useful e.g. for producing
 enzymes with altered properties, is based on cyclic extension of unpaired
 primers.

PS Disclosure; SEQ ID NO 2; 481pp; English.

CC The present invention describes a method for the recombination of nucleic
 CC acid (NA) templates. The method comprises: (a) providing at least two
 CC double-stranded templates (T1, T2) having different 5' and 3' regions in
 CC their sense strands; (b) contacting the templates with at least one each
 CC of forward and reverse primers that anneal, respectively, only to the 3'-
 CC region of the antisense strand of T1 and to the 3'-region of the sense
 CC strand of T2; (c) extending the primers by no more than 1000 nucleotides
 CC (nt); (d) separating extended primers from their templates; (e)
 CC reannealing the extended primers, to either template; and (f) repeating
 CC cycles of steps (b)-(e) until at least one full-length extension product,
 CC i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
 CC method starting from at least one each of antisense and sense single-
 CC stranded templates; (2) generating a recombined polypeptide (1) with
 CC altered properties by expressing recombination products of the new
 CC processes and screening the polypeptides formed for properties different
 CC from those of polypeptides from either of the templates; and (3) (1)
 CC produced by method (2). The methods are useful for the recombination of
 CC nucleic acid templates of interest, and generation of a recombined
 CC polypeptide having altered properties. The method can be used to make
 CC combined genes that express proteins e.g. enzymes, cytokines, growth
 CC factors, viral proteins or microbial antigens with altered properties,
 CC e.g. stability, activity or specificity. The present sequence is used in
 CC the exemplification of the present invention.

XX Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 7; Length 716;

Best Local Similarity 84.0%; Pred. No. 4.5e-300;

Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAFSLDLSIATSKXNGXNGXNAXXASXXXXXXGSGXLPPTXXXTOLDIYEXXAD 60
 Db 1 MAFSLDLSIATSKXNGXNGXNAXXASXXXXXXGSGXLPPTXXXTOLDIYEXXAD 60
 QY 61 PXTDDXXELDGLTIGDVGAARKGRXRVXVDSDEIRKXIDKSVFRLXOLXNSVYGV 120
 Db 55 P-TDSTLELDGYSINLDVVSARKGRPVKVDSDISIKIDKSVFRLXOLXNSVYGV 113
 QY 121 TGFSGSADRTTEDAISLOKALEHOLCGVLPXTSDSFXLGRLENSLPLEVVRGAMTRV 180
 Db 114 TGFSGSADRTTEDAISLOKALEHOLCGVLPSSPFRGRLNSLPLEVVRGAMTRV 173
 QY 181 NSLTRGSAVRVYVLEALTNFLNHTGTPVPLRGITISASGDLSPSYIAAATGHPDSKV 240
 Db 174 NSLTRGSAVRVYVLEALTNFLNHTGTPVPLRGITISASGDLSPSYIAAATGHPDSKV 233
 QY 241 HVHHEGEEKIMKARELIAFGLEPVYLGKXEGELGVNGTAVASAMATALDAMLSLS 300
 Db 234 HVHHEGEEKILVAREMAFLNLEPVYLGKXEGELGVNGTAVASAMATALDAMLSLS 293
 QY 301 QALTALTVAMVGHAGSFPHLDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEVKVD 360
 Db 294 QSLTALTVMVGHAGSFPHLDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEVKVD 353
 QY 361 DEGLIQDRYPLRTSQNGPLVSDMITAHAVLSLBAQOSTTNDPLIVENKXTHHGNGF 420
 Db 354 DEGLIQDRYPLRTSQNGPLVSDMITAHAVLSLBAQOSTTNDPLIVENKXTHHGNGF 413
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLEMLNAGNNGRGPCLAAEDPSLSYHCKGDDIA 480
 Db 414 QAAAVANTWEKTRILGAIQKLNFTQLEMLNAGNNGRGPCLAAEDPSLSYHCKGDDIA 473
 QY 481 AAAYSELGLHNPVTTTHVQAPAMGNAVNSIALISARTRAEANDVSLLIATLTYCLQ 540

Db 474 AAAYSELGLHNPVTTTHVQAPAMGNAVNSIALISARTRAEANDVSLLIATLTYCLQ 533
 QY 541 AVDLRAMEFEFKQFPXXXXLXOHFGXXXTXXXXXXELXKXVXXKLKLEQNTSYDL 600
 Db 534 AILRLAIEFFKQFPALVSLIDQHFGSAMGNSLRDELVEKNKTLAKRLKEQNTSYDL 593
 QY 601 EPRWDAFSAKGTVTXLSXXXXXVSLAANAMKVAAEKAISLTXXVXXFXAPSS 660
 Db 594 VPRWDAFSAKGTVTXLS--TSLAANAMKVAAEKAISLTQVRETFWMSAAT 650
 QY 661 SSPALXYSPTRVLSFVREELGVKARGDVFELGQETIGSNVSRITYEAKSGRIYXV 720
 Db 651 SSPALXYSPTRVLIYAFREELGVKARGDVFELGQETIGSNVSKITYEAKSGRIYXV 710
 QY 721 LVXKLA 726
 Db 711 LKXKLA 716

RESULT 10

ID AAE20668 standard; protein; 716 AA.

XX AAE20668;

DT 01-JUL-2002 (first entry)

DE R. glutinis phenylalanine ammonia lyase (PAL) mutant, EPI8tm-6.

XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome P-450; cytochrome P-450 reductase; liquid crystal polymer;
 KW ICP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst; mutant; variant; mutcin.

XX Rhodotorula glutinis.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 540 /note= "Wild-type Ile substituted with Thr"

XX WO200210407-A1.

XX 07-FEB-2002.

PF 23-JAN-2001; 2001WO-US002099.

PR 27-JUL-2000; 2000US-00627216.

PA (DUPO) DU PONT DE NEMOURS & CO I.

PI Gatenby AA, Sariastani FS, Tang X, Qi WW, Vannelli T;

DR WPI, 2002-280635/32.

DR N-PSDB; AAD33100.

PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT polypeptide, or mutant TAL polypeptide, useful for producing
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.

PS Disclosure; Page; 139pp; English.

CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the

CC Plant cytochrome p-450 and the cytochrome p-450 reductase into *E. coli*.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used
 CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC mutant PAL. Note: This sequence is not shown in the specification.
 CC however it is constructed based on the PAL enzyme shown as SEQ.ID.NO:8
 CC (AAE20653) in the sequence listing

XX Sequence 716 AA;

Query Match 89.7%; Score 3019; DB 5; Length 716;

Best Local Similarity 84.0%; Pred. No. 7, 2e-300; Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

```

QY 1 MAPSLDSIATSYANGXNGXHAAXXASXXXXXXXAXGSLPTTXXXTOLDIVEXXIAD 60
DB 1 MAPSLDSISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQYTDIVEXKMLAA 54
QY 61 PXTDDXHELDGYSILTIGDYVGAARKGRXRVYXDSDEIRKXIDKSYEFIRKXOLXNSVYGV 120
DB 55 P-TDSTLELDGYSILNIGDYVSAARKGRPVYKXDSDEIRSKIDKSYEFIRSOQLSMSVYGV 113
QY 121 TGFSGSADRTEDATSLQKALLEHQLCGVLPFSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTEDATSLQKALLEHQLCGVLPFSXDSFXLGRGLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEBALTNFLNHGITPIVPRGTISAGSLSPDLYTAAITGHPDSKV 240
DB 174 NSLTRGSAVRLVLEBALTNFLNHGITPIVPRGTISAGSLSPDLYTAAITGHPDSKV 233
QY 241 HVNHEGEEKIMAREIALFGLEPVYLGKREGGLVNGVAVASMATLALDANHLSTLS 300
DB 234 HVNHEGEEKILVAREMALFNLEPVYLGKREGGLVNGVAVASMATLALDANHLSTLS 293
QY 301 QATLALTEAMVGHAGSFHFLHDVTRPHPTQIEVARNIRTLLEGSGXPAVHNEEVKXD 360
DB 294 QSTLALTEAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLDESGRFAVHNEEVKXD 353
QY 361 DEGLIQDRPLRTSPQMGPLVSDMIAHAAVLSLEAGOSTDNLIVENKXTHGNGF 420
DB 354 DEGLIQDRPLRTSPQMGPLVSDMIAHAAVLSLEAGOSTDNLIVENKXTHGNGF 413
QY 421 QASAVNTEKTRTALATLIGKINFTQLEMLNAGMNRG-PSCLAEDPSLSYHCKGLDIA 480
DB 414 QASAVNTEKTRTALATLIGKINFTQLEMLNAGMNRG-PSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVCYLQ 540
DB 474 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVCYLQ 533
QY 541 AVDLRAMEPEFKQKQFPKXXLLXOHFGKXXTXXXXXXEIXKKYKXIKRLEQNSVDL 600
DB 534 AIDLRTAEPEFKQKQFPALVSLIDQHFSGAMTGSMLRDELVEKNVKTAKLEQNSVDL 593
QY 601 EPRMHAFAFATGTVEKLSXXXXXXVSLAANAMKVAAEKALSLTRXVXXEFKXAPSS 660
DB 594 VPRMHAFAFATGTVEKLS---TSLSLAANAMKVAAEKALSLTRXVXXEFKXAPSS 650
QY 661 SSPALXLSRRTVYVSFVREELGVKARGDVFLGKQEVTTISNSRLEYAKISGRINXV 720
DB 651 SSPALXLSRRTVYVSFVREELGVKARGDVFLGKQEVTTISNSRLEYAKISGRINNV 710
QY 721 LVKMLA 726
DB 711 LKMLA 716

```

RESULT 11
AAE20654

ID AAE20654 standard; protein; 716 AA.
 XX
 XX AAE20654;
 AC
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant.
 XX
 XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 KW LCP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst; mutant; variant; mutcin.
 XX
 OS Rhodotorula glutinis.
 OS Synthetic.
 OS
 PN WO200210407-A1.
 PD 07-FEB-2002.
 PF 23-JAN-2001; 2001WO-US002099.
 PR 27-JUL-2000; 2000US-00627216.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI Gatebby AA, Sarislanı FS, Tang X, Qi W, Vannelli T;
 DR WPI; 2002-280635/32.
 DR N-PSDB; AAD33076.
 XX
 PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT polypeptide, or mutant TAL polypeptide, useful for producing
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.
 PS
 XX Claim 2; Page 98-101; 139pp; English.
 CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into *Escherichia coli* underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
 CC plant cytochrome p-450 and the cytochrome p-450 reductase into *E. coli*.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used
 CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC PAL mutant
 XX
 SQ Sequence 716 AA;

Query Match 89.7%; Score 3019; DB 5; Length 716;
 Best Local Similarity 84.0%; Pred. No. 7, 2e-300; Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

```

QY 1 MAPSLDSIATSYANGXNGXHAAXXASXXXXXXXAXGSLPTTXXXTOLDIVEXXIAD 60
DB 1 MAPSLDSISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQYTDIVEXKMLAA 54
QY 61 PXTDDXHELDGYSILTIGDYVGAARKGRXRVYXDSDEIRKXIDKSYEFIRKXOLXNSVYGV 120
DB 55 P-TDSTLELDGYSILNIGDYVSAARKGRPVYKXDSDEIRSKIDKSYEFIRSOQLSMSVYGV 113
QY 121 TGFSGSADRTEDATSLQKALLEHQLCGVLPFSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

```

```

Db      114 TGFSGSADRTEDDAISLQKALLHQLCGVLPSSFDRLGRGJENSLPEVVRGAMTIRV 173
Qy      181 NSLTRGHSAYRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 240
Db      174 NSLTRGHSAYRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 233
Qy      241 HVHEGEXKIMXAREALFGLPEVYLGPKEGGLVNGTAVASVMTALAHAMLSLS 300
Db      234 HVHEGEXKILYAREALFNLPEVVLGPKEGGLVNGTAVASVMTALAHAMLSLS 293
Qy      301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGSXPAVHHEEVKVD 360
Db      294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGSXPAVHHEEVKVD 353
Qy      361 DEGLIRQDRYPLRTSPQWIGPLVSDMTHAAVLSLISAGSTTDNPLIDVENKXTHGNGF 420
Db      354 DEGLIRQDRYPLRTSPQWIGPLVSDMTHAAVLSLISAGSTTDNPLIDVENKXTHGNGF 413
Qy      421 QASAVNTMEKTRILALIGKLNFTQTEMNAGNNGRLPSCIAEDPESLTHCKGLDIA 480
Db      414 QAAVANTMEKTRILALIGKLNFTQTEMNAGNNGRLPSCIAEDPESLTHCKGLDIA 473
Qy      481 AAAYTSELGHANPYTTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLATLHYCVLQ 540
Db      474 AAAYTSELGHANPYTTHVQPAEMNGQAVNSLALISARRTESNDVLSLLATLHYCVLQ 533
Qy      541 AVDLRAMPEFFKQFPKXXKXLLKHQFGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db      534 AIDRATATEFFKQFGAIVLSLIDQHFSGAMTGSNRLDELVEKXNKTAKRLSEQNSYDL 593
Qy      601 EPRWHDASXATGTVVEVXLSKXXXXXVSLAAVNAVMAVAAEKALSLTRXVXXFXVXAPS 660
Db      594 VPRWHDASFAAGTVVEVLS--TSLSLAAVNAVMAVAAESALSTRQVRETFWASAAT 650
Qy      661 SSPALXLSPTRTLYLFFVREELGYKARQGVFLGQVETLIGSNRYIETAKSRINXV 720
Db      651 SSPALXLSPTRTLYLFFVREELGYKARQGVFLGQVETLIGSNRYIETAKSRINXV 710
Qy      721 LVKMLA 726
Db      711 LKMLA 716

RESULT 12
AAE16390
ID AAE16390 standard; protein; 716 AA.
AC AAE16390;
XX
DT 09-APR-2002 (first entry)
XX
DE Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL).
XX
KW Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW aerospace application; chemical; red yeast; mutant; mutcin.
XX
OS Rhodotorula glutinis.
OS Synthetic.
XX
FT Key Location/Qualifiers
FT Misc-difference 540 /note= "Wild type Ile substituted with Thr"
XX
XX WO20011071-A2.
XX
XX 15-FEB-2001.
XX
XX 03-AUG-2000; 2000MO-US021156.
XX
XX 06-AUG-1999; 99US-0147719P.

```

```

XX (DUPO) DU POINT DE NEMOURS & CO E. I.
XX Tang X, Vannelli TM, Qi WW, Sariafiani S, Gatenby AA;
XX WPI; 2002-121549/16.
XX N-PSDB; AAD26924.
XX
XX Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
XX comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia-
XX -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
XX activity.
XX
XX Claim 14; Page 68-70; 75pp; English.
XX
XX The invention relates to a method for producing para-hydroxycinnamic acid
XX (PHCA). The method comprises converting cinnamate to PHCA by converting
XX glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
XX route or by generating a new biocatalyst possessing enhanced tyrosine
XX ammonia-lyase (TAL) activity. The method is useful for the biological
XX production of PHCA which is useful as a monomer for the production of
XX liquid crystal polymers (LCP), where the LCP is used in electronic
XX connectors, telecommunications and aerospace applications. LCP is also
XX useful in medical devices, as well as chemical, and food packing
XX applications due to its resistance to sterilizing radiation. The present
XX sequence is red yeast, Rhodotorula glutinis (Rhodospiridium toruloides)
XX mutant PAL derived by substituting Ile at position 540 with Thr. Mutant
XX PAL enzyme possesses enhanced TAL activity
XX
XX Sequence 716 AA:
XX
XX Query Match 89.7%; Score 3019; DB 5; Length 716;
XX Best Local Similarity 84.0%; Pred. No. 7.2e-300;
XX Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;
Qy 1 MAPSLDITSTXANXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXX 60
Db 1 MAPSLDITSTXANXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXXNXX 54
Qy 61 PXTDDXELDGYSLTLDGVYGAARKGRXVYXDSDEIRKXIDKSVLEFLRXQJXNSVYGV 120
Db 55 P-TDSTLELDGYSLNLDGVYSAARKGRPVYVXDSEIRKIDKSVLEFLRQJXNSVYGV 113
Qy 121 TGFSGSADRTEDDAISLQKALLHQLCGVLPSSFDRLGRGJENSLPEVVRGAMTIRV 180
Db 114 TGFSGSADRTEDDAISLQKALLHQLCGVLPSSFDRLGRGJENSLPEVVRGAMTIRV 173
Qy 181 NSLTRGHSAYRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 240
Db 174 NSLTRGHSAYRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 233
Qy 241 HVHEGEXKIMXAREALFGLPEVYLGPKEGGLVNGTAVASVMTALAHAMLSLS 300
Db 234 HVHEGEXKILYAREALFNLPEVVLGPKEGGLVNGTAVASVMTALAHAMLSLS 293
Qy 301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGSXPAVHHEEVKVD 360
Db 294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGSXPAVHHEEVKVD 353
Qy 361 DEGLIRQDRYPLRTSPQWIGPLVSDMTHAAVLSLISAGSTTDNPLIDVENKXTHGNGF 420
Db 354 DEGLIRQDRYPLRTSPQWIGPLVSDMTHAAVLSLISAGSTTDNPLIDVENKXTHGNGF 413
Qy 421 QASAVNTMEKTRILALIGKLNFTQTEMNAGNNGRLPSCIAEDPESLTHCKGLDIA 480
Db 414 QAAVANTMEKTRILALIGKLNFTQTEMNAGNNGRLPSCIAEDPESLTHCKGLDIA 473
Qy 481 AAAYTSELGHANPYTTHVQPAEMNGQAVNSLALISARRTAEANDVLSLLATLHYCVLQ 540
Db 474 AAAYTSELGHANPYTTHVQPAEMNGQAVNSLALISARRTESNDVLSLLATLHYCVLQ 533
Qy 541 AVDLRAMPEFFKQFPKXXKXLLKHQFGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600

```

[illegible]

XX 26-FEB-2003; 2003MO-US005708.
 XX 26-FEB-2002; 2002US-0360279P.
 XX (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA Milano J, Tang X;
 PI WPI; 2003-721766/68.
 DR N-PSDB; ADE36055.
 XX Recombination of nucleic acid templates, useful e.g. for producing
 PT enzymes with altered properties, is based on cyclic extension of unpaired
 PT primers.
 PS Disclosure; SEQ ID NO 4; 481bp; English.
 XX The present invention describes a method for the recombination of nucleic
 CC acid (NA) templates. The method comprises: (a) providing at least two
 CC double-stranded templates (T1, T2) having different 5' and 3' regions in
 CC their sense strands; (b) contacting the templates with at least one each
 CC of forward and reverse primers that anneal, respectively, only to the 3'-
 CC region of the antisense strand of T1 and to the 3'-region of the sense
 CC strand of T2; (c) extending the primers by no more than 1000 nucleotides
 CC (nt); (d) separating the primers from their templates; (e)
 CC reannealing the extended primers, to either template; and (f) repeating
 CC cycles of steps (b)-(e) until at least one full-length extension product,
 CC i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
 CC method starting from at least one each of antisense and sense single-
 CC stranded templates; (2) generating a recombinant polypeptide (1) with
 CC altered properties by expressing recombination products of the new
 CC processes and screening the polypeptides formed for properties different
 CC from those of polypeptides from either of the templates; and (3) (1)
 CC produced by method (2). The methods are useful for the recombination of
 CC nucleic acid templates of interest, and generation of a recombinant
 CC polypeptide having altered properties. The method can be used to make
 CC combined genes that express proteins e.g. enzymes, cytokines, growth
 CC factors, viral proteins or microbial antigens with altered properties,
 CC e.g. stability, activity or specificity. The present sequence is used in
 CC the exemplification of the present invention.
 XX Sequence 716 AA;

Query Match 89.7%; Score 3019; DB 7; Length 716;
 Best Local Similarity 84.0%; Pred. No. 7.2e-300;
 Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MABSLDISIATXGKXNGXNHAAXXAXXXXXXAXXGXYLPTTXXTOLDIVEXXLAD 60
 DB 1 MABSLDISIATXGKXNGXNHAAXXAXXXXXXAXXGXYLPTTXXTOLDIVEXXLAD 54
 QY 61 PXTDDXXELDGYSLTIGDVGAARKGRXVRVXDSEIRAKIDKSYEPRLRXQXNSVYCVT 120
 DB 55 P-IDSTLELDGYSLNIGDVGAARKGRXVRVXDSEIRAKIDKSYEPRLRXQXNSVYCVT 113
 QY 121 TGGGASADTETDPAISIQKXLEHQCGLPTXSXDSFKXGRLXNSLPLEVVRGAMTRV 180
 DB 114 TGGGASADTETDPAISIQKXLEHQCGLPTXSXDSFKXGRLXNSLPLEVVRGAMTRV 173
 QY 181 NSLTRGSAVRLVLEALTFNLNHTPIVPLRGTTISAGDSPISTYIAAATGHPDSKV 240
 DB 174 NSLTRGSAVRLVLEALTFNLNHTPIVPLRGTTISAGDSPISTYIAAATGHPDSKV 233
 QY 241 HVXHEGXEKTMAREALALFGLBPVVLGPEXGLGVNGTAVASAMTALDHAHLISLS 300
 DB 234 HVXHEGXEKTMAREALALFGLBPVVLGPEXGLGVNGTAVASAMTALDHAHLISLS 293
 QY 301 QALTALVTEAMVAGAGFHPFLHDVTRPPTQTEVARNIRTLXGSAFVAHHEBEVKYD 360
 DB 294 QSLTAMTEAMVAGAGFHPFLHDVTRPPTQTEVARNIRTLXGSAFVAHHEBEVKYD 353
 QY 361 DEGLIRDRYPLRTSPQMLGPLVSDMIHAHVALISEAGSSTTDNPLIDVENKXTHHGNF 420

DB 354 DEGLIRDRYPLRTSPQMLGPLVSDMIHAHVALISEAGSSTTDNPLIDVENKXTHHGNF 413
 QY 421 QASAVXNTWEKTRALALIGLNTOLTETMNAQNGSLPSCLAEPSPSTHYCKGLDIA 480
 DB 414 QAAAVANTMEKTRGLAOLIGLNTOLTETMNAQNGSLPSCLAEPSPSTHYCKGLDIA 473
 QY 481 AAATSEIGHLANPTTVOJPAEMGNOAVNSIALISARPTAEANDVLSLALATLYCVLQ 540
 DB 474 AAATSEIGHLANPTTVOJPAEMGNOAVNSIALISARPTAEANDVLSLALATLYCVLQ 533
 QY 541 AVDLRAMEFEFKQCPYXXXXLXQHPGXKXTXXXXXELXXKXXKXLRLEQNSYDL 600
 DB 534 AIDLRAFEFEFKQCPYXXXXLXQHPGXKXTXXXXXELXXKXXKXLRLEQNSYDL 593
 QY 601 EPRMHDATSKATGTVEVLESLSSXXVSLAAVNAWTVAXAEKASITLFRYRXFWXAPSS 660
 DB 594 VPRMHDATSKATGTVEVLESLSSXXVSLAAVNAWTVAXAEKASITLFRYRXFWXAPSS 650
 QY 661 SSPALXYLSPTRVLYSFVREELGVKARQGVFLGKQEVITIGSNVSRIZAKISGRINXY 720
 DB 651 SSPALXYLSPTRVLYSFVREELGVKARQGVFLGKQEVITIGSNVSRIZAKISGRINXY 710
 QY 721 LYXMLA 726
 DB 711 LYXMLA 716

RESULT 15

AA881099
 ID AAP81099 standard; protein; 716 AA.

XX AAP81099;
 AC 12-DEC-1990 (first entry)

DT 12-DEC-1990 (first entry)

DE Sequence of Rhodospiridium toruloides IFO 559 cells L-phenylalanine

DE ammonia lyase (PAL).

XX Aspartame; artificial sweetener; enzyme.

OS Rhodospiridium toruloides.

PN BP260919-A.

XX 23-MAR-1988.

PF 15-SEP-1987; 87EP-00308119.

PR 16-SEP-1986; 86JP-00215864.

PA (MITR) MITSUI TOKAI CHEM INC.

PI Fukuhara N, Yoshino S, Yamamoto K, Se T, Sone S, Nakajima Y;

PI Suzuki M, Makiyuchi N;

XX WPI; 1988-079223/12.

DR N-PSDB; AA881399.

PT New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for

PS Claim 1; Page 21-23; 37pp; English.

CC PAL is useful in prodn. of L-phenylalanine from cinnamic acid and an NH3

CC donor (process claimed). L-phen is an essential AA and starting material

CC for sweeter aspartame. E.coli transformants will produce PAL without

CC induction by expensive AAs; no surfactant is needed to increase cell wall

XX permeability

XX Sequence 716 AA;

Query Match 89.7%; Score 3018; DB 1; Length 716;

Best Local Similarity 83.9%; Pred. No. 9,1e-300;
Matches 609; Conservative 26; Mismatches 81; Indels 10; Gaps 3;

```
QY 1 MAPSLDSIATSYXNGXNGXHAAXXAXXXXXXXXAGSGPPTXXQTOLDIVEXXAD 60
Db 1 MAPSLDSISHSPANGVASAKQAVNGAS-----TNLAVAGSHLPTQTQVDIVEXKLLAA 54
QY 61 EPTDDXELDGLSLGLGVYGAARKGRVYVXDSDEIRXKIDKSEFELRXOLXNVYGV 120
Db 55 P-TDSTLELDGSLNLDGVVSAARKGRVYVXDSDEIRSKIDKSEFELRSQLSMVGVT 113
QY 121 TGFSGADRTEDATSLQKALLEHQLCGVLPYSXDSFXLGRLENSLPLEVVGAMTIRV 180
Db 114 TGFSGADRTEDATSLQKALLEHQLCGVLPSSPDSFRLGRLENSLPLEVVGAMTIRV 173
QY 181 NSITRGHSAVRLVYVLPALTNELNHGITTPIVLRGTTISAGDLSPLSTYAAITGHPDSKY 240
Db 174 NSITRGHSAVRLVYVLPALTNELNHGITTPIVLRGTTISAGDLSPLSTYAAITGHPDSKY 233
QY 241 HXHEGEXKIMKAREALAFGLPEVYLGKREGGLVNGTAVSASWATLALHDAMLSLS 300
Db 234 HXHEGEXKILYAREAMALFNLPEVYLGKREGGLVNGTAVSASWATLALHDAMLSLS 293
QY 301 QALTALVEAMVGHAGSFHPLHDVTRPHPTQIEVANIRITLLEGSXFAVHHEEVKXD 360
Db 294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLLEGSFPAVHHEEQVKXD 353
QY 361 DEGITLQDQYPRRTSPQWIGPLVSDMHAHAYLSLQAGOSTTDNPLIDVENKXTHGNGF 420
Db 354 DEGITLQDQYPRRTSPQWIGPLVSDMHAHAYLTLEAGOSTTDNPLIDVENKXTHGNGF 413
QY 421 QASAVXNTEKTRALALIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRGLAQIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHANPVTTHVQPAKNGQAVNSLALISARRTAEANDVLSLLATHLYCYLQ 540
Db 474 AAAYTSELGHANPVTTHVQPAKNGQAVNSLALISARRTAESNDVLSLLATHLYCYLQ 533
QY 541 AVDLRAMEPEFFKQFQFXXXXLXOHFGXXXTXXXXXXELXXXKVKXLKRLQETNSYDL 600
Db 534 AIDLRAIEPEFFKQFQFPAVLSLIDQHFGSAMTGSNLRDELVEKVKMLAKRLQETNSYDL 593
QY 601 EPRMDAFSXATGTVEVEXLSXXXXXVSLAAVANAMKVAAXAEKALSTRXVXXFXKAPSS 660
Db 594 VPRMDAFSFAAGTVEVLS--TSLSLAAVANAMKVAABESALSLTRQVRETFWASAAT 650
QY 661 SSPALXYLSRPTVUXSFVREELGVYARGGVFLGKQEVTTGSNYSRIYEAIKSGRINXV 720
Db 651 SSPALXYLSRPTVILYAFVREELGVYARGGVFLGKQEVTTGSNYSRIYEAIKSGRINNV 710
QY 721 LVHQLA 726
Db 711 LKQMLA 716
```

Search completed: September 9, 2004, 10:04:03
Job time : 92.3693 secs




```

Db 354 DEGLIRORRYPLRTSPQWLGVLSDLIHAHVAVLTEAGOSTDNDPNDVENTSHHGNGF 413
Qy 421 QASAVNTMEKTRTALALIGLNTQLTQTEMNAGNRGRLPSCLAADPSLSYHCKGLDIA 480
Db 414 QAAVANVTMEKTRGLGLAIGLNTQLTQTEMNAGNRGRLPSCLAADPSLSYHCKGLDIA 473
Qy 481 AAATSEIGLANPVTTHVQPAEMNQAANSIALISARPTAENDVLSLLATHTHYCVLQ 540
Db 474 AAATSEIGLANPVTTHVQPAEMNQAANSIALISARPTTESNDVLSLLATHTHYCVLQ 533
Qy 541 AVDRAMEFEFKQFPKXXXXLXQHGXXXXXXELXXKVKXKLKRLKRLKRLKRLKRLKRL 600
Db 534 AIDRAIEFEFKQFPKPAIVSLIDQHGSAWTSNLRDELVEKVNKTAKRLKRLKRLKRLKRL 593
Qy 601 EPRWMDAFSXAATGTVEXLSXXXXXVSLAAVNAWMAAEEKALSLTRKVRXXFWXAPSS 660
Db 594 VPRWMDAFSFAAGTVEXLS--TSLSLAAVNAWMAAEEKALSLTRKVRRTFWSAAS 650
Qy 661 SSPALXLSPTRTVLYSFVREELGVKARQDVFVGKQEVTTGSNVSRIYEAIKGRINXV 720
Db 651 SSPALXLSPTRTVLYSFVREELGVKARQDVFVGKQEVTTGSNVSRIYEAIKGRINNV 710
Qy 721 LVKMLA 726
Db 711 LVKMLA 716

```

RESULT 2

```

A:Species: Rhodospiridium toruloides
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
R:Accession: A29607
R:Author: U.G.; Gilbert, H.J.; Oram, J.D.; Minton, N.P.
Gene 58; 189-199, 1987
A:Title: Complete nucleotide sequence of the Rhodospiridium toruloides gene coding for p
A:Reference number: A29607; MUID:88112870; PMID:2828184
A:Accession: A29607
A:Molecule type: DNA
A:Residues: 1-693 <ANS>
A:Cross-references: GB:W18261; NID:G169745; PIDN:AAA33883.1; PID:G169746
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:188-190/Cross-link: 5-imidazoliumone (Ala-Gly) #status predicted
F:189/Modified site: dehydroalanine (Ser) #status predicted

```

Query Match 87.7%; Score 2951; DB 2; Length 693;

Best Local Similarity 86.4%; Pred. No. 2, 4e-224;

Matches 590; Conservative 23; Mismatches 66; Indels 4; Gaps 2;

```

Qy 44 PTTXXQLDIVEXXLADPXTDXXELDGYSLTLGVGAAKGRKXVYXDSDEIRKXIDK 103
Db 15 PTTQVQVIVDEVMIAAP-TDSTLELDGSLNLGVVSAARKGRVRYXDSDEIRKXIDK 73
Qy 104 SVEFLXQLXNSVYGVTTGFGSADTRTDAISLQKALLEHQLGCVLPTXSDSFXLGRGL 163
Db 74 SVEFLXQLXNSVYGVTTGFGSADTRTDAISLQKALLEHQLGCVLPTXSDSFXLGRGL 133
Qy 164 ENSLPJEVVGAMTIVNSLTRGSAVPLVLEALTNLNGITPIYVPLRTGISASDLS 223
Db 134 ENSLPJEVVGAMTIVNSLTRGSAVPLVLEALTNLNGITPIYVPLRTGISASDLS 193
Qy 224 PLSYIAAATGHDPSKVHYHEGKXIKXAREALFGLBEPVLGPKXGLVNGTAVSA 283
Db 194 PLSYIAAATGHDPSKVHYHEGKXIKXAREALFGLBEPVLGPKXGLVNGTAVSA 253
Qy 284 SVATLALHDAHMLSLISQALTLTYEAVVGAAGSHPLHVTTPHPTQLEVNANTITL 343
Db 254 SVATLALHDAHMLSLISQALTLTYEAVVGAAGSHPLHVTTPHPTQLEVNANTITL 313
Qy 344 EGSXFAVHHEEYKVDDEGLRDRYPLRTSPQWLGVLSDMIHAHVAVLSLEAGOSTTD 403

```

```

Db 314 EGSXFAVHHEEYKVDDEGLRDRYPLRTSPQWLGVLSDMIHAHVAVLSLEAGOSTTD 373
Qy 404 NPLIDVENKCKTHHGGFOASAVXNTMEKTRTALALIGLNTQLTQTEMNAGNRGRLPSC 463
Db 374 NPLIDVENKCKTHHGGFOASAVXNTMEKTRTALALIGLNTQLTQTEMNAGNRGRLPSC 433
Qy 464 AADPSLSYHCKGLDIAAAVYSELGLANPVTTHVQPAEMNQAANSIALISARPTAEA 523
Db 434 AADPSLSYHCKGLDIAAAVYSELGLANPVTTHVQPAEMNQAANSIALISARPTES 493
Qy 524 NDVLSLLATHTHYCVLQAVDLAMEFEFKQFPKXXXXLXQHGXXXXXXELXXK 583
Db 494 NDVLSLLATHTHYCVLQAVDLAMEFEFKQFPKXXXXLXQHGXXXXXXELXXK 553
Qy 584 VXXKLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRL 643
Db 554 VXXKLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRLKRL 610
Qy 644 ISLTRVXXKXFXADSSSPALXVYSPRTVLYSFVREELGVKARQDVFVGKQEVTTGS 703
Db 611 ISLTRVXXKXFXADSSSPALXVYSPRTVLYSFVREELGVKARQDVFVGKQEVTTGS 670
Qy 704 NVSRIYEAIKGRINXVLVKMLA 726
Db 671 NVSRIYEAIKGRINNVLVKMLA 693

```

RESULT 3

```

A:Species: Rhodotorula rubra
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
R:Accession: S01999; S01990
R:Filipula, D.; Vasilec, C.A.; Levy, A.; Sykes, A.; Strausberg, R.L.
Nucleic Acids Res. 16, 11381, 1988
A:Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula
A:Reference number: S01999; MUID:89083582; PMID:3205749
A:Accession: S01999
A:Molecule type: DNA
A:Residues: 1-713 <FII>
A:Cross-references: EMBL:X13094; NID:G3284; PIDN:CAA1486.1; PID:G295942
R:Vasilec, C.A.; Strausberg, R.L.; Sykes, A.; Levy, A.; Filipula, D.
Nucleic Acids Res. 16, 11382, 1988
A:Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal ge
A:Reference number: S01990; MUID:89083583; PMID:3205750
A:Accession: S01990
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-674 <VAS>
A:Cross-references: EMBL:X13095
C:Genetics:
A:Introns: 110/2; 139/2; 372/3; 428/2; 482/3
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:217-219/Cross-link: 5-imidazoliumone (Ala-Gly) #status predicted
F:218/Modified site: dehydroalanine (Ser) #status predicted

```

Query Match 86.8%; Score 2922.5; DB 2; Length 713;

Best Local Similarity 82.0%; Pred. No. 4, 4e-222;

Matches 595; Conservative 33; Mismatches 85; Indels 13; Gaps 5;

```

Qy 1 MAPSDISATSGXANGXNGXAXXXXXXAXXAXXAXXAXXAXXAXXAXXAXXAXXAXXAXX 60
Db 1 MAPSDISATSGXANGXNGXAXXXXXXAXXAXXAXXAXXAXXAXXAXXAXXAXXAXXAXX 59
Qy 61 PXTDXXELDGYSLTLGVGAAKGRKXVYXDSDEIRKXIDPSVFLXQLXNSVYGVTT 120
Db 60 AGATDQIRLDGYSLTLGVGAAKGRKXVYXDSDEIRKXIDPSVFLXQLXNSVYGVTT 119
Qy 121 TFGGSADTRTDAISLQKALLEHQLGCVLPTXSDSFXLGRGLNSLPJEVVGAMTIV 180
Db 120 TFGGSADTRTDAISLQKALLEHQLGCVLPTXSDSFXLGRGLNSLPJEVVGAMTIV 179

```


QY 181 NSLTRGSAARIYVLEATLTPLNHGTPITPIYPIGTTISASGDSLPLSTYAAITGHPSKY 240
 Db 180 NSLTRGSAARIYVLEATLTPLNHGTPITPIYPIGTTISASGDSLPLSTYAAITGHPSKY 239
 QY 241 HVHGEKEXIMXAFEAIALPGLSEPVYLGPKXGLGVNGTAVASAMATIALHDAHMLSLLS 300
 Db 240 HV--DG--KIMSAGDEALTKGLQGVVLGPKXGGLGVNGTAVASAMATIALTDAAVLSILA 295
 QY 301 QALTALIVEAMVNGAGSFPHPLHDVTRPHRPOLEVAANITTLLEGXYFAVHNHEEVYVKD 360
 Db 296 QALTALIVEAMVNGAGSFPHPLHDVTRPHRPOLEVAANITTLLEGSKYAAVHNHEEVYVKD 355
 QY 361 DEGLIRDDRYPLRTSPOMLGPIVSDMTHAAVLSLEAGOSTTDPLIDVENKXYHNGNF 420
 Db 356 DEGLIRDDRYPLRTSPOMLGPIVSDMTHAAVLSLEAGOSTTDPLIDVENKXYHNGNF 415
 QY 421 QASAVXNTMTXTRIALALLIGKANTQUTLTMNAGMNGFL9SCLAEDPSSLSYHKGJLDA 480
 Db 416 MASSVGNTMTXTRIALAVLMGVRSPFTQUTLTMNAGMNGFL9SCLAEDPSSLSYHKGJLDA 475
 QY 481 AAATYSELGHANVTVTHVOPAEKNGQAVNSIALISARRTEANDVLSLLATLHYCYLQ 540
 Db 476 AAATYSELGHANVTVTHVOPAEKNGQAVNSIALISARRTEANDVLSLLATLHYCYLQ 535
 QY 541 AVDLRAFEFEKQFQFXXXXXLXQHFQXXXXTXXXXXELXXXRVYXXLXKRLBQTSYDL 600
 Db 536 AVDLRAFEFEHTKFAFEPVTELLKQHFQALAT-----AEVEDKVRKKSIVYKRLQQNSYDL 590
 QY 601 EPRHNDFAFSXATGVTVEXLSXXXXXVSIAVNMKVAAXEKAISLRVYVXXXFWXAPSS 660
 Db 591 EQRHNDHFSVATGVTVEXL--QEVSLASLNMKVAACEKAIALTRSVRDSFWAAPSS 647
 QY 661 SSPALXYLSPETRYLVSFVREBELGVKARQGVFLGXQGVTTIGSNVSSIYEAIKSGRINXY 720
 Db 648 SSPALXYLSPETRYLVSFVREBELGVKARQGVFLGXQGVTTIGSNVSSIYEAIKSGCIAPV 707
 QY 721 LVKXMA 726
 Db 708 LVKXMA 713

RESULT 4
 G84787
 Phenylalanine ammonia-lyase (PAL) [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: G84787
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M., Koo, H.; Moffat, K.S.; Cronin, L.A.; Sherman, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Accession: G84787
 A:Accession: G84787
 A:Molecule type: DNA
 A:Residues: 1-725 <SIO>
 A:Cross-references: GB:AA002093; NID:G6596547; PIDN:AA018156.2; GSPDB:GN00139
 C:Genetics:
 A:Map position: 2
 C:Superfamily: histeridine ammonia-lyase

Query Match 28.8%; Score 969; DB 2; Length 725;
 Best Local Similarity 35.5%; Fred. No. 3,3e-68;
 Matches 258; Conservative 120; Mismatches 257; Indels 92; Gaps 21;

QY 40 GSXLPPTXXTQDIVEXXLADPYTDDXXELDGYSLTLGDVVGAAARKGRXRVXDSDEIRX 99
 Db 39 GAAAEQMGSHLDEYKRVAR-FRKPVNVNGEITLTIQVAAISITGNSVVELSETARA 97
 QY 100 KIDKSVFLRQLXNSV--YGYTTGFGSADRTREDAISLOQALLHQLCGVLP7SKXSF 157
 Db 99 KIDKSVFLRQLXNSV--YGYTTGFGSADRTREDAISLOQALLHQLCGVLP7SKXSF 157

Db 96 GVNASDWMESNMKGTDSYGVTTGFGATSRRTKNGVALQKELIRPLNAGIFGSETET - 156

Qy 158 XLRGRLNLSPLFVVRGAMTIRNLSIRGSAVNLVLEALTNLNGIRPIVLRGTIS 217

Db 157 -----SHLPFSATRAAMLVIRINTLLQGFEGFIRLELLEALNNNNITPSLBRGTTT 210

Qy 218 AGDGLSPSYIAAATTTGHPDSKRVHHEGSKIKMXAREALAPGLEB--VVLGKEGLGL 275

Db 211 ASGDLVPLSYIAGLTGRPNKA--TGPRNG--EALTAEAFKLAGISGCFDLOKEGAL 267

Qy 276 VNGTAVASVATLALHDHMTLSLSQALTALTVAAMGHAGSEFPFLADTRPHPTQIEV 335

Db 268 VNGTAVGSGMASMVFETNNVLSVLAELISVAFVAMSGKP-EFDDHLLTHRIKHPGQIEA 326

Qy 336 AANNRTILEGSXF---AVHHEEVKVDDEGILRDRYPLRTSPQMLGLVSDMIHAHA 391

Db 327 AALMEHIIDSSYMTLQKLEHMDLPQK-----KQDRYALRTSPQMLGQIE--VIRYA 379

Qy 392 VLSLENG-QSTNDNPLDVENKXTHHGNFQASAVXNTEKTRIALALLIGKLTQLTLEM 450

Db 380 TKSIEREINISVNDNPLDIVSRNRKAIHGNFGGTPIGVSMNTRIALAIGRLMFAQPSSEL 439

Qy 451 LNAAGNRGLPSTCL-AAEDPSLSYHCKGDDIAAAAYTSLGLANPVYTHVQAPMGQAQV 509

Db 440 VNDPFINNGPLPSULTKSRPSPLDYGFKGALMASYCSLEQLANPVTHSHQSAQHQDV 499

Qy 510 NSIALISARTAEANDVLSLLATHLYCVTLQAVDLRAMEFEFKQEPXXXXLXHQFGX 569

Db 500 NSLGLISRSKSTSEAVDILKLMSTPLVALCOAVDLRHEENLRLQ-----543

Qy 570 KTXKXXXXXELKKXV-----KKLXKKRL--EQNTSYDLEPRMHDFAFSXAT 612

Db 544 --TVKNVTVSQVAKKVLTTGVNGELHPSPFCERDLKRVDRQVYTVADP-----CSAT 595

Qy 613 GTVEXLSSXXXXXYSVLAAVNA-----KVAVAEKRAIS--LTRYVXXFWXAPS 659

Db 596 YPLQLQKLRQ---VYDHALINGESEKNAVTSLPHKIGAFEBELKAVLPKEVEBARAAAYDN 652

Qy 660 SSSPALXLYSP-RTKVLVLSFYREELGVAKRGDVFGLQGEYTISSNRSRIYEAKSKSRIN 718

Db 653 GTSALPNRIKESRYPLEFRVFEELGT-----ELLTGKVTSPGESEFDKVTALCEGKII 707

Qy 719 XVLVKKYL 725

Db 708 DPMMECL 714

RESULT 5

SS2990

phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999

C:Accession: SS2990; PQ0140

R:Wanner, L.A.; Li, G.; Ware, D.; Somsich, I.E.; Davis, K.R.

A:Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.

A:Reference number: SS2990; NCBI:95195160; PMID:7888622

A:Accession: SS2990

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-725 <NAN>

A:Cross-references: EMBL:L33677; NID:9497418; PIDN:AAC18870.1; PID:9497419

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994

R:Ohl, S.; Hedrick, S.A.; Chory, J.; Lamb, C.J.

Plant Cell 2, 837-848, 1990

A:Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsis

A:Reference number: PQ0140; NCBI:93005677; PMID:2152131

A:Accession: PQ0140

A:Molecule type: DNA

A:Residues: 1-240 <OHL>

A:Cross-references: GB:X62747; NID:916408; PIDN:CAA44609.1; PID:916409

A:Experimental source: Ecotype Columbia

C:Comment: Phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to ti

C:Genetics:
 A:Gene: PAL1
 A:Introns: 143/2
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 F:211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:112/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.7%; Score 965; DB 2; Length 725;
 Best Local Similarity 35.5%; Pred. No. 6,9e-68;
 Matches 258; Conservative 119; Mismatches 258; Indels 92; Gaps 21;

```

QY 40 GSXLPPTXXTQDLIVEXXLDADPTDDXXELDYSILTGVDV--VGAARKGXVRYXSDDEIRX 99
DB 39 GAALAEQKMGSHLDEVKMAVE--FRKPVNVLGGSTLITGVAAIISTIGNSVKVLESTARA 97
QY 100 KIDKSVFELRXQDLXNV--YGVTTGFGSADRTEDAIISLOKALLEHOLCGVLPFSXDSF 157
DB 98 GVNASDPMVMESNMKGTDSYGVTTGFGATSHRTKXGVALQKELRFLNAGIFSGTKET- 156
QY 158 XLGGLGELNSLPLEVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGTTTPVLRGTIS 217
DB 157 -----SHLPHSATRAAMLVRLNTLLQGSGRFELIATISFLNNITPSPLRGTIT 210
QY 218 ASGDLSPSTIAAATGHPDSKVHXXHEGKIMKAREALFGLER--VVLGPRGGL 275
DB 211 ASGDLVPLSYIAGLITGRPKSA--TGPNQ--EALTEBAFKLAGISSGFPLQREGAL 267
QY 276 VNGTAVASAVATLALHDAMLSTLSQALTALVTEAMVHAGSFHFLHDVTRPHTQIEV 335
DB 268 VNGTAVSGASAVLLEETNLVLAELISAVFAEWSGKP--EFTDLTHRLKHPGQIEA 326
QY 336 ARNRTLEGGXSF-----AVHHEEVKVKDDEGLRQDRYPLRTSPWGLPLVSDMTHA 391
DB 327 AAVMEHLIDSSSYMKLAQKLEHNDPLQKP-----KODRYALRTSPQWLGPOIE--VIRYA 379
QY 392 VLSLEAG--QSTTDNPLIDVENKXTHHGNFQASAVXNTEKTRLALIGKLNFTOLTEN 450
DB 380 TKSIEREINSVNDNPLIDVSRNKAHGGNFQGPICGVSMDNTRLAIRAIIGKLMFAQFSEL 439
QY 451 LNAAGNRGLPSCA--AEDPSLSYHCKGLDIAAAATSELGHLANVTTHVQPAEMGNOAV 509
DB 440 VNDPVNNGPNSNLTARNSPLDGFGEAETIAMSYSSELQYLANVTSHVQSAQOHNDV 499
QY 510 NSLALISARRTAEANDVLSLLATHLVYCVLAQVLDLRAHEFEFKQFPXKXLLXQHFQX 569
DB 500 NSLGLISSRKTSADVILKLMSTTFLVAICQAVDLRHEENLRQ----- 543
QY 570 XXTXXXXXXELXXKXV-----XXKLKKL--EQTNSYDLEPRMHAFAFXAT 612
DB 544 --TVKQTVSQVAKVAVLTGNGELHPSRCECDLLKVDVREQVYIADDP-----CSAT 595
QY 613 GTVEXLSXXXXXXVSLAAVNA-----MKVAXAEKAIIS--LTRVYKXXFWXAPS 659
DB 596 YPLIQKLRQ---YIVDHALVNGSEKNATSTIFHKIGAFEBELKAVLPKEVEAARAAYDN 652
QY 660 SSSPALXYLSP--RTRVLVSFVRBELGVKARGDVPFLGKQVTTGSNVSRIYEAIKSGRI 718
DB 653 GTSALPNRIKECRSYPLRYFVEBELGT-----ELTGEKVTSPGEEFKVPTAICEGKI 707
QY 719 XVLVKKL 725
DB 708 DPMVECL 714

```

RESULT 6
 506475
 phenylalanine ammonia-lyase (EC 4.3.1.5) - rice
 C:Species: Oryza sativa (rice)
 C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 21-May-1999
 C:Accession: S06475; A33246; J00156
 R:Minami, E.; Ozeki, Y.; Matsuoaka, M.; Koizuka, N.; Tanaka, Y.
 Eur. J. Biochem. 185, 19-25, 1989

A>Title: Structure and some characterization of the gene for phenylalanine ammonia-lyase
 A:Reference number: J00156; KWID:90032676; PMID:2806257
 A:Accession: S06475
 A:Molecule type: DNA
 A:Residues: 1-701 <MIN>
 A:Cross-references: EMBL:X16099
 A>Note: the authors translated the codon GTC for residue 66 as Leu, CAG for residue 35
 F:189-191/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:190/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.5%; Score 958.5; DB 2; Length 701;
 Best Local Similarity 36.0%; Pred. No. 2.1e-67;
 Matches 255; Conservative 116; Mismatches 266; Indels 71; Gaps 21;

```

QY 40 GSXLPPTXXTQDLIVEXXLDADPTDDXXELDYSILTGVDV--VGAARKGXVRYXSDDEIRX 97
DB 16 GAALAEQKMGSHLDEVKMAVQ--FREPLVKGATLRVQVAAVQAQDAARVAVEIDEEA 74
QY 98 EKXIDKSVFELRXQDLXNV--SVYGVTTGFGSADRTEDAIISLOKALLEHOLCGVLPFSX 155
DB 75 RRPVKAASWITTCIAHGGDIYGVTTGFGGSHRTKDGPLQVELLRHNAIGFSGSD 134
QY 156 SFXLGGLGELNSLPLEVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGTTTPVLRGT 215
DB 135 G-----HTLPSTVAAMLVRLNTLLQGSGRFELIATISFLNNITPSPLRGTIT 186
QY 216 ISAGDLSPSTIAAATGHPDSKVHXXHEGKIMKAREALFGLER--VVLGPRGGL 273
DB 187 ITASGDLVPLSYIAGLITGRPKSA--ISPDGRK--VDAAEFKLAGIEGGEFTLNPEGL 243
QY 274 GLVNGTAVASAVATLALHDAMLSTLSQALTALVTEAMVHAGSFHFLHDVTRPHTQIEV 333
DB 244 AIVNGTSVGSALAAVVMFDANILAVSEVLASVCEVMNGKPEYTDHLTKL--KHNGST 302
QY 334 EVARNIRTLBGSXAVHHEEVKVKDDEGL--RQRYPLRTSPQWLGPOIE--VIRYA 391
DB 303 DAAATMEHILAGSSSPMSHAK--KYNEWDPLKCKQDRVALRTSPQWLGPOI--QVIRA-A 357
QY 392 VLSLEAG--QSTTDNPLIDVENKXTHHGNFQASAVXNTEKTRLALIGKLNFTOLTEN 450
DB 358 TKSIEREYNSVNDNPLIDVHGGKALHGGNFQGTIGVSDNARLALANIGKLMFAQFSEL 417
QY 451 LNAAGNRGLPSCA--AEDPSLSYHCKGLDIAAAATSELGHLANVTTHVQPAEMGNOAV 509
DB 418 VNEFYNNGLTNSLNGSRPSIDYGFKEITIMASYSSELQYLANPITNHVQSAQOHNDV 477
QY 510 NSLALISARRTAEANDVLSLLATHLVYCVLAQVLDLRAHEFEFKQFPXKXLLXQHFQX 569
DB 478 NSLGLVSARKTLEADVILKLMSTTFLVAICQAVDLRHEENKISSVKNCTQYAKKVLTM 537
QY 570 XXTXXXXXXELXXKXKXKLXRLDQNTSYDLEP--RHHDAFSXATGVVE--XLSXXXXXX 626
DB 538 NPTGLSSARFSEKNTLAIADR--EAVFSYADDPESANYPIMQKLRVAVLVEHALTSGRRR 596
QY 627 VSLAAVNAWKVAXAEKAIISLTRVYKXXFWXAPSSSSP-----ALVYLS 670
DB 597 RGLRVLDHQQVGG-----APLCAPAGDRGRPRRRORTAPYANRIVES 640
QY 671 RTRVLVSFVRBELGVKARGDVPFLGKQVTTGSNVSRIYEAIKSGRI 717
DB 641 RSPFYLRVREBELG-----VFLTGEKLTSPGEECKXKVLGISQKTL 682

```

RESULT 7
 T09777
 phenylalanine ammonia-lyase (EC 4.3.1.5) - loblolly pine


```

Db      384 NPLIDVSRNKAHIGNFQGTPIGSMNDTRLAIAIGKLMFAQFSELVNDFFNNGLPENL 443
Qy      464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPTTHVQPAEMGNQAVSLISARRTAE 522
      444 SGRNPSLDYGFAGBIAMASYCSELQFLANPTVTHVQSAEHNQDVAVSLGISRRKISE 503
Qy      523 ANDVLSLLATLHYCYLQAVDLRAMEFEFKQFPYXXXLLXQHFE-----GXXTXXXXX 577
      504 AVEILKLMSTFTLVGLCOALIDRLHEENLKSTVKNTVSSVAKRVLTJMGVNGELHPSRFE 563
Qy      578 XEIXXXKXXK-----LXKRLBOT-----NSYDLEPRWHDASFXYATGV 616
      564 KDLRLRVDRERFAYITDPCSATYPLMOKLRQTLVHALKNDNERMNTSTIQKIALTFE 623
Qy      617 EXLSSXXXXXVSLAAVNAKVAAXEALISLTXVXXXFXAPSSSPALXYLSP----- 670
      624 DEK-----ALLPKEVESAR-----AALLESNPAL-----PNRIEBC 655
Qy      671 RTVLVSFYREELGVKARGDVFLGKQETIGSNVSRIVEATKSGINXVLVYKML 725
      656 RSYPLVKFVKELGT-----EYLTGEKVTSPGEERFVKFIAMSKGEIIDPLECL 705

```

RESULT 11

```

S04463
phenylalanine ammonia-lyase (EC 4.3.1.5) - parsley
C/Species: Petroselinum crispum (parsley)
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-May-1999
C/Accession: S04463
R/Lot: R.; Dietrich, A.; Hahlbrock, K.; Schulz, W.
EMBO J. 8, 1641-1648, 1989
A/Title: A phenylalanine ammonia-lyase gene from parsley: structure, regulation and ider
A/Reference number: S04463; M0ID:89356635; PMID:2767049
A/Molecule type: DNA
A/Residues: 1-716 <LOI>
A/Cross-references: EMBL:X15473
C/Genetics:
A/Intons: 136/2
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/202-204/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F/203/Modified site: dehydroalanine (Ser) #status predicted

```

```

Query Match      28.1%; Score 946.5; DB 2; Length 716;
Best Local Similarity 36.7%; Pred. No. 1,9e-66;
Matches 254; Conservative 110; Mismatches 286; Indels 43; Gaps 18;

```

```

Qy      49 TQDIVEXXLADPXTDDXXELDGYSLTGADVVG-AARRGRVRYXSDSEIRXKIDKSYEF 107
      40 SHLDEVKQWVAE-YRKPVVKLGGETLLISQVAIAISARDGSGVTVELSEAAAGVAKASDW 98
      108 LRKQLNSV--YGVTTGFGGSADRTEDALISOKLLHOLCGVLPTXKDFXGRGLEN 165
      99 VWDSPMKGTDSTGVTTGFATSHRRKQGGALQKLIFFLNAGI-----FNGSDN 149
Qy      166 SLELEVRAQMTIRVNSLTRGSAVALVLEALTFNLHGITPIVPLRGITISASGDSL 225
      150 TLPHSATRAMLVRIHTLLQYSGIRFEITLAIETFLMONTPTCLPRGTTASGDLVPL 209
Qy      226 SYIAAALITGHPSKXVYHHEGKEXIMKAREALALGLER--VLGPKGGLGVNGTAVSA 283
      210 SYIAGLLTRRPNSKA-VGPTGV--ILSPERAKLAGVGSGFEELQPKGLALVNGTAVGS 266
Qy      284 SMATLALHDAHMLSLISQALTLATYEAMVGHASFPRLADVTARHPPTQIEVANIRITL 343
      267 GNASWMLFEENILAVLAEMSAIFAEVQGR-EFTDHLTHLKHHPQDIAAIMEHIL 325
Qy      344 EGSXPAVHHEEVKVKODEGILRQRYPLRTSPQMLGRVSDMIHAAYLSLEAGSTTD 403
      326 DGSAY-VKAQCKLHEHNDPLQKPKODRYALRTSPQMLGPQI-EVIRSTKMTIEREINSND 383
Qy      404 NPLIDVENKCTHNGCFQASAVXNMEXKTRLALALIGLNTFQULTENMAEMNGLSCL 463

```

```

Db      384 NPLIDVSRNKAHIGNFQGTPIGSMNDTRLAIAIGKLMFAQFSELVNDFFNNGLPENL 443
Qy      464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPTTHVQPAEMGNQAVSLISARRTAE 522
      444 SGRNPSLDYGFAGBIAMASYCSELQFLANPTVTHVQSAEHNQDVAVSLGISRRKISE 503
Qy      523 ANDVLSLLATLHYCYLQAVDLRAMEFEFKQFPYXXXLLXQHFGXXXXXXXXELXX 582
      504 AVEILKLMSTFTLVGLCOALIDRLHEENLKSTVKNTVSSVAKRVLTJMGVNGELHPSRFE 563
Qy      583 KYXKILXKLEQNTSYDLEP--RMHDAFSXATGVYVELSSXXXXXVSLAAVNAKVAAXA 640
      564 KDLRLRVDR-EYIFAYITDPCSATYPLMOKLRQTLVHALKNDNERMNTSTIQKIALTFE 622
Qy      641 EKALIS--LTRVXXXFXWAPSSSPALXYLSP-----RTVLVSFYREELGVKARGDV 692
      623 EDELKALPLKEVESA-PALESNPAL-----PNRIEERSYPLVKFVKELGT-----EY 672
Qy      693 FLGKEVTTGSNVSRIVEATKSGINXVLVYKML 725
      673 LTGEKVTSPGEERFVKFIAMSKGEIIDPLECL 705

```

RESULT 12

```

S22891
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - soybean
C/Species: Glycine max (soybean)
C/Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C/Accession: S22891
R/Frame: R.L.; Vodka, L.O.
DNA Seq. 1, 335-346, 1991
A/Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max.
A/Reference number: S22891; M0ID:92190550; PMID:1799692
A/Accession: S22891
A/Molecule type: DNA
A/Residues: 1-713 <FRA>
A/Cross-references: EMBL:X52953; NID:g18376; PID:CA37129.1; PID:g18377
C/Genetics:
A/Gene: PHL1
A/Intons: 131/2
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/199-201/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F/200/Modified site: dehydroalanine (Ser) #status predicted

```

```

Query Match      28.1%; Score 946; DB 2; Length 713;
Best Local Similarity 36.6%; Pred. No. 2.1e-66;
Matches 271; Conservative 105; Mismatches 293; Indels 72; Gaps 23;

```

```

Qy      14 NGXNGXHAAXXAXSXXXXXXAXAGSXLPTXXXTQDIVEXXLADPXTDDXXELDQYS 73
      5 NGHQSGFCLSTAKGNNDPLMWGAA---EAMGSHLDEVKRYVAE-YRKPVVRIGGET 59
      74 LTLGDVYGAARGRYRYXSDSEIRXKIDKSEVFLRXQLNSV--YGVTTGFGGSADRT 131
      60 LTIQVAAVAGDHGVAVELSESAREGVKASEWMNNGTDSYGVTTGFGATSHRR 119
Qy      132 EDAISLOKALLHOLCGVLPTXSDFXLGRGLENS--LPLVVGAMTIRVNSLTRGSA 189
      120 KGGALQKELIRFLNAGI-----FNGTSHHTLPHTATAAALVRIHTLLQGYSG 170
Qy      190 VRLVLEALTFNLHGITPIVPLRGITISASGDSLPSYIAAALITGHPSKXVYHHEGEXK 249
      171 IFEITLAIETKLNNVNPCLDLRTIITASGDLVPLSYIAGLLTGRPSKA-VGPGS-E 227
Qy      250 IIXAERIALFEL--EPVVLGPKBGLGVNGTAVASASATLALHDAHMLSLISQALALT 307
      228 VLNAREAEELASINSEFELQPKGGLAVNGTAVGSGLASVWLFENAILAVLSVAIF 287
Qy      308 VAMVGHASFPRLADVTARHPPTQIEVANIRITLLEGSXPAVA---HEEEVYVKXDEG 363
      288 AEVQGR-EFTDHLTHLKHHPQDIEAALIMEHILDOSSYMKAAKXHLHIDPLQKP--- 343

```

QY 364 ILSDRYPLATSPOMIGPLVSDMIIHAVALSLAEG-QSTTDPLIDVENKXTHGGNFOA 422
 Db 344 --KODRYALATSPOMIGPLIE--VIRFSTKSIEREINSVNDPLIDVSRNKALHGGNFOG 399
 QY 423 SAVANTWEKTRIALALIGKLNFTQLTETMLNAGNRGLPSC-L-AAEDPSLSYHCGLDIAA 481
 Db 400 TPIGVSMNDTRLLALASIGKLMFAQPSFELVNDPFYNNGLPSNLTPASRNPGLDYGKCAETAM 459
 QY 482 AATSEIGHLANPVTTHVQPAEMGNQAVNSLALISARTRAEANDVLSLLATHLVYCYQA 541
 Db 460 ASYCELSQYLANPVTTHVQSAEQHNDVNSLIGLISRKTEALIELIKLMSSTFLALCOA 519
 QY 542 VDIAPAEFFERKQFPXPKXXLLXQHFQXXXLXKXXKXKXKRL--EQTNSYD 599
 Db 520 IDLRHLEENLKNVKNVTSQVSKR--ILLTGNGELHPSRCEKDLKVDREYIFSYI 576
 QY 600 LEPAMDPAFSXATGTVEXISSXXXXVSLAANA-----KVAANAERKAIS--L 646
 Db 577 DDP-----CSATYPLMQKLRQ--VLVDHALVNAECEKDVNSIFQKIALFEEELKNL 627
 QY 647 TRYVRXXFXKAPSSSPAL--XYLSPRTVLVSFVSEELGVKARQGVFLGKQEVYISGN 704
 Db 628 PKEVEGA-RAAYESGKRAIPNKIQEGRSYPLKFRFEEELGT-----GLITGEKVRSPGE 681
 QY 705 VSRIYEAIKSGRINXVLVKML 725
 Db 682 FDKLFAMCOGKIIDPMECL 702

RESULT 13

S66343
 phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco
 C/Species: Nicotiana tabacum (common tobacco)
 C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
 C/Accession: S66343
 R/Pukawska-Akeda, T.; Kung, S.; Watson, J.C.
 Plant Mol. Biol. 30, 711-722, 1996
 A/Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicotiana
 A/Reference number: S66343; MUID:96194455; PMID:8624404
 A/Accession: S66343
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-715 <PDB>
 A/Cross-references: EMBL:M64466; NID:G170349; PDB:AAA3122.1; PID:G170350
 C/Genetics:
 A/Gene: PAL1
 A/Introns: 133/2
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/201-203/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/204/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.0%; Score 943; DB 2; Length 715;
 Best Local Similarity 36.4%; Pred. No. 3.7e-66;
 Matches 253; Conservative 107; Mismatches 254; Indels 82; Gaps 18;
 QY 68 ELQGVSTLGDVVGAA---RKGXRVVXSDERKXIDKSVFLKXQLKNSV--YGVTG 122
 Db 53 KLGGESLTVAGVAAIAVRDASANGVVEEERKAGVKSASDWVDSMKKGTDSYGVTG 112
 QY 123 FGGSADRTEDAIASLQKALEHQLCGVLPFSXDSFLXGRLNS--LPLEVVRGAMTIRV 180
 Db 113 FGATSHRRTKNGALQGEILRFPMAGV-----FGKGTESHTLPHSATRAAMLVRI 163
 QY 181 NSLTRGSAVRLVLTNLFNLHGTTPLVPLRGITISAGDLSPISTYIAAATGHPDSKV 240
 Db 164 NTLQGYSGIRFELLEITPLINSNTPLRGITITAGDILVPISYIAGLTGHPNSKA 223
 QY 241 HVHGEKEXIKARAEALFGLRP--VLGPKEGGLGVGTAVSASMTALADHAMLSL 298
 Db 224 -VGENG--ETLNEEARVAGVAGVFELQPKEGGLAVNGTAVGSMAVMVLFDSNIIAV 280

QY 299 LSQLATLVEAMVGHAGSPFPLHDVYTRPPTQIEVARNIRTLBEGSXPAVHEEYKV 358
 Db 281 MSEVLSAIFABVMNKP-EFTDHLTHKXHPGQIEAALIMETHLDDSSY-VKAAQCLHE 338
 QY 359 KODEGILRODRPPLTSPQMLGPLVSDMIIHAVALSLAEGSTTDNPLIDVENKXTHGG 418
 Db 339 MDLPQPKODRYALATSPOMIGPLQ--EVIRATQIEREINSVNDPLIDVSRNKALHGG 397
 QY 419 NFOASAVXWTEKTRIALALIGKLNFTQLTETMLNAGNRGLPSC-L-AAEDPSLSYHCGL 477
 Db 398 NFOGTPIGSMNALLALASIGKLMFAQPSFELVNDYNNGLPSNLTPASRNPGLDYGKCA 457
 QY 478 DTAAPATSEIGHLANPVTTHVQPAEMGNQAVNSLALISARTRAEANDVLSLLATHLVY 537
 Db 458 EIAMASYCELSQYLANPVTTHVQSAEQHNDVNSLIGLISARKTEALIELIKLMSSTFLA 517
 QY 538 VLQAVDLAMEFEFRKQFPXPKXXLLXQHF-----GXXXTXXXXXXEXKXKXKX 587
 Db 518 LQALIDLRLHLEENLKNVKNVTSQVAKETLTMGANGELHPARFCKELLAIVDREYIFAY 577
 QY 588 -----LXKRLQOT-----NSYDLERMDPAFSXATGTVEXISSXXXXVSLA 631
 Db 578 ADDPGSCNYPFLMQKLRQVLVDHANNGSEKVNNSIFQKIGARED-----ELKA 627
 QY 632 VNAKVAANAERKALISLTVVRXXFXKAPSSSPAL--XYLSPRTVLVSFVSEELGVKAR 689
 Db 628 VLPKVESRPAAL-----BSGNPAIPNRITCEKRSYPLKFRFEEELGT----- 669
 QY 690 GDFVLGKQEVYISGNVSRIVEAIKSGRINXVLVKML 725
 Db 670 -ELITGEKVRSPGECDKVFYFAMCGKIIDPMECL 704

RESULT 14

T46172
 phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
 N/Alternate names: protein T4D2.190
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
 C/Accession: T46172; S52391
 R/Makatura, G.; Fartmann, B.; Dauner, D.; Seerr, W.; Holland, R.; Weichselgartner, M.
 submitted to the Protein Sequence Database, December 1999
 A/Reference number: 223025
 A/Accession: T46172
 A/Molecule type: DNA
 A/Residues: 1-717 <NYS>
 A/Cross-references: EMBL:AL12958; NID:G6434245; PDB:CA864229.1; PID:G6630746
 A/Experimental source: cultivar Columbia; BAC clone T4D2
 R/Wanner, L.A.; Li, G.; Ware, D.; Somselich, I.E.; Davis, K.R.
 Plant Mol. Biol. 27, 327-338, 1995
 A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
 A/Reference number: S52390; MUID:95195160; PMID:7868622
 A/Accession: S52391
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-205, 'H', 207-234, 'E', 236-717 <MAN>
 A/Cross-references: EMBL:U33678; NID:G497420; PDB:AAU18871.1; PID:G497421
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C/Genetics:
 A/Gene: PAL2
 A/Map position: 3
 A/Introns: 135/2
 A/Note: T4D2.190
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/203-205/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/204/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.9%; Score 939.5; DB 2; Length 717;
 Best Local Similarity 34.2%; Pred. No. 6.9e-66;
 Matches 251; Conservative 122; Mismatches 262; Indels 99; Gaps 20;
 QY 49 TQLDIVEXKLLADPXT-----DDXKE-----LDGYSITLGDVVGAA 84

```

Db      15  TKAAVTTKTLADPLNGLAADQKSGSHLDEYVKWEEYRPPVNLGGELLTIGQVAIAST 74
Qy      85  KGRVRYXDSDEIRXXIXIDKSEVEFLRXQLXNSV--YGVTTFGSGSADRTEDAISLQKALL 142
Db      75  VGSQVKVELAETSRAGVYKASSDWVWESNMKGTDSGVTTTGFATSHRRHKGTALQTELI 134
Qy      143  EHQLCGVLPSSXSFYKLGKLENSLLEVRGAMTRVNSLTRGSAVRLVLELITNPL 202
Db      135  RFLNAGIFGNTKTC-----HTLPQSATRDAMLVRVNTLLQYSGIGRIFELLEITSL 187
Qy      203  NHGITFVPRGTISASGDLSPASYTAAITGHPDSKVNHHXEGEKIMXAREALRGL 262
Db      188  NHMISPLPRGTTTSGDLVPLSTYAGLLTGRPNKA-TGPDG--ESTLKEAEKXGI 244
Qy      263  EP--VVLGPRGEGLVNTRVSAVSAMATLALHDAHMLSLISQALLTVAWAGHAGSHP 320
Db      245  STGFEPDLPQEGALVNGTAVSGVMSAVLFRANQAVLAEVLISAFVAVNSGKP-EFTD 303
Qy      321  FLADVTREPHTQTEVARNIRTLLEGXSF-----AVHHEEVKVDDEGLRQDRPLRTS 375
Db      304  HLHRLKHPGQLEAAIMEHILDSYKLAQKTHENDPLQK-----KODRYALRTS 357
Qy      376  PQWLGPLVSDMHAHVAVLSLEAG-QSTTDNPLIDVENKXTHHGNFQASAVXNTMEKTRL 434
Db      358  PQWLGQPIR--VIRQATKSIEREINSVNDNPLIDVSRNKALHGNFQGPFGVSMNTRL 415
Qy      435  ALALICKLNTQUTETMLNAGNRGLPSC--AAEDPSLSYHCGDLIAAAATSEIGHAN 493
Db      416  ALAIAIKLMPAQSELVNDFYNNNGPNSNLTASSNSLDYGFQAGAIAMASYCSELQYLAN 475
Qy      494  PVTTHVQPEMGNQAVNSLALISARITAEANDVLSLATHLYCYLQAVDIRAMEFEFK 553
Db      476  PVTSHVQSAHQNHQDONSIGLISRTSSEAVDILKLMSTFLVGIQAVDIRHLEENLRQ 535
Qy      554  QEPKXXXXLIXHFGKXXXIXXXXXXELKKV-----XXKLXKRL-EQTN 596
Db      536  -----TVKNTVSQVAKKVLTTGNGELHPSRCEKDLKVVDRQVF 577
Qy      597  SYDLEP--RWHDAFSXATGTVEVEXLSSXXXXVSLAAVAMKVAEKAIS--LTRXRX 652
Db      578  TYVDDPCSATPELMQRLROYIVDHALSNGETEKNAVTSIFQKIGAFEEELKAVLPKEVEA 637
Qy      653  XFWXADSSSPALXYLSP-RTRVLYSFVREELGVKARAGDVFLEKQEVTTIGSNVRIYE 711
Db      638  ARAAYNGTAFLIPNRKICEGRSYPLRFVREELGT-----LITGEKVSPSEEDKVF 692
Qy      712  IKSGRINXVLVXKL 725
Db      693  MCEGKLIDPLMDCL 706

RESULT 15
T03663
C/Species: Nicotiana glauca (Common tobacco)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
R/Accession: T03663
A/Accession: T03663
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-712 <PEL>
A/Cross-references: EMBL:X78269; NID:G633596; PIDN:CA55075.1; PID:G633597
A/Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicid
A/Experimental source: cultivar Samsum NN; tissue-type leaves after infection by tobacco
C/Genetics:
A/Gen: PAL
C/Function:
A/Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid

```

```

C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/198-200/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F/199/Modified site: dehydroalanine (Ser) #status predicted

Query Match      27.9%; Score 938; DB 2; Length 712;
Best Local Similarity 35.2%; Pred. No. 9e-66;
Matches 253; Conservative 115; Mismatches 260; Indels 90; Gaps 18;

Qy      49  TQDIYEXKLADPXTDDXKLELDYSLTLDGVYGAARK--GRVRYXDSDEIRXKIDKVE 106
Db      33  SHLDEKKNVAB-FRRPVYKLGELLTYAQVALAKONAKTVKELSEGARAGYKASSD 91
Qy      107  FLRXQLXNSV--YGVTTFGSGSADRTEDAISLQKALLEHQLCGVLPSSXSFYKGRGLE 164
Db      92  VWDSMSKGTDSGYVTTGFGATSHRRTKGALQKELRFLNAGV-----FGNGTE 142
Qy      165  --NSLPLEVRGAMTRVNSLTRGSAVRLVLELITNPLNHGTRPIVPRGTISASGDL 222
Db      143  SCHTLPQSGTRAAMLVRINTLLQYSGIRFELLEITLKLHNVTPLCLPLGTTTASDL 202
Qy      223  SPLSYTAAAITGHPDSKVNHHXEGEKIMXAREALALFGLP--VVLGPRGEGLVNGTA 280
Db      203  VPLSYTAGLLTGRPNKA-IGPNC--ETLNAEAFRVAGVNSGFEELQPKGALVNGTA 259
Qy      281  VASAVATLALHDAHMLSLISQALLTVAWAGHAGSHPFLHVTREPHTQIEVARNIR 340
Db      260  VSGLSAVLFPANTLAFSEVLSAIFAEVNSGKP-EFTDHLTKLKHHPQIEAALTE 318
Qy      341  TLLEGSPAVHHEEVKVDDEGLRQDRPLRTSPQWLGPLVSDMHAHVAVLSLEAGOS 400
Db      319  HILDGSSY-VKARQKLHETDPLQKQDRYALRTSPQWLGQIEVIRATGMIEREINS 376
Qy      401  TIDNPLIDVENKXTHHGNFQASAVXNTMEKTRLALIGKLNFTQUTETMLNAGNRGLP 460
Db      377  VNDNPLIDVSRNKALHGNFQGPFGVSMNARLALSIGLMPAQSELVNDFYNNGLP 436
Qy      461  SCI-AAEDPSLSYHCGDLIAAAATSEIGHANPVTTHVQPEMGNQAVNSLALISAR 519
Db      437  SNLTGRNNSLDYGRKGEIIMASYCSELQFLANPVNHHVQSAHQNHQDONSIGLISRK 496
Qy      520  TAEANDVLSLATHLYCYLQAVDIRAMEFEFKQEPKXXXXLIXHFGKXXXIXXXXXE 579
Db      497  TAEAVDILKLMSTYVALCAIDIRHLEENLRN-----A 531
Qy      580  LXXKXKXKXKXLEQTNVYDLEP-----RWHDAFSXATGTVEVEXLSS 621
Db      532  VKNTVSQVAKKRTLMGANGELHPSRCEKDLRVVDRBYFRVADACSAVPLMQKLRQ 591
Qy      622  XXXXVSLAAVAMKVAAXA-----EKAISLTRVYXXFWXAPSSSPALXYLSP----- 670
Db      592  ---VLVDHALENGGENEKANNSIFQKILAFEGELKAVLPKEVESARISLENGNPAIANRI 648
Qy      671  ---FRRVLYSFVREELGVKARAGDVFLEKQEVTTIGSNVRIYEAIKSGRINXVLVXKL 725
Db      649  KCGRSYPLRFVREELG-----AELLTGEKVRSPGEEDKVFMTMCGOITDLSLECL 701

```

Search completed: September 9, 2004, 10:08:48
 Job time : 38.1203 secs

